เครื่องมือการตลาดเชิงนวัตกรรม

โดยใช้เหมืองความคิดเห็นบนไมโครบล็อก

นายวิลาส จ่ำเลิศวัฒน์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรดุษฎีบัณฑิต

สาขาวิชาธุรกิจเทคโนโลยีและการจัดการนวัตกรรม (สหสาขาวิชา)

บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย

ปีการศึกษา 2554

INNOVATIVE MARKETING TOOL

BY APPLYING OPINION MINING ON THE MICRO-BLOG

Mr.Wilas Chamlertwat

A Dissertation Submitted in Partial Fulfillment of the Requirements

for the Degree of Doctor of Philosophy Program in Technopreneurship and Innovation Management (Interdisciplinary Program)

Graduate School

Chulalongkorn University

Academic Year 2011

Thesis Title	INNOVATIVE MARKETING TOOL BY APPLYING OPINION MINING ON THE MICRO-BLOG
By	Mr.Wilas Chamlertwat
Field of Study	Technopreneurship and Innovation Management
Thesis Advisor	Assistant Professor Pattarasinee Bhattarakosol, Ph.D.
Thesis Co-advisor	Tippakorn Rungkasiri, Ph.D.

Accepted by the Graduate School, Chulalongkorn University in Partial Fulfillment of the Requirements for the Doctoral Degree

…… Dean of the Graduate School
(Associate Professor Pornpote Piumsomboon, Ph.D.)

THESIS COMMITTEE

…….. Chairman
(Associate Professor Supawan Tantayanon, Ph.D.)

…….. Thesis Advisor
(Assistant Professor Pattarasinee Bhattarakosol, Ph.D.)

…….. Thesis Co-advisor
(Tippakorn Rungkasiri, Ph.D.)

…….. Examiner
(Assistant Professor Pongpun Anuntavoranich, Ph.D.)

…….. Examiner
(Supawin Watcharamul, Ph.D.)

…….. External Examiner
(Choochart Haruechaiyasak, Ph.D.)

วิลาส ฉ่ำเลิศวัฒน์ : เครื่องมือการตลาดเชิงนวัตกรรมโดยใช้เหมืองความคิดเห็นบนไมโครบล็อก. (INNOVATIVE MARKETING TOOL BY APPLYING OPINION MINING ON THE MICRO-BLOG) อ. ที่ปรึกษาวิทยานิพนธ์หลัก : ผศ.ดร.ภัทรสินี ภัทรโกศล, อ. ที่ปรึกษาวิทยานิพนธ์ร่วม : ดร.ทิพภากร รังคสิริ, 114 หน้า.

วิธีการวิจัยตลาดแบบดั้งเดิมอาจจะไม่เพียงพอสำหรับการแข่งขันในธุรกิจปัจจุบันเพราะความต้องการเวลาและทรัพยากรที่สูง ดังนั้นนักวิจัยจึงขอนำเสนอทางเลือกที่เรียกว่า “ระบบการวิเคราะห์ความรู้สึกบนไมโครบล็อก” ที่พัฒนาจากหลักการการทำเหมืองความคิดเห็นเพื่อวิเคราะห์ความคิดเห็นของผู้บริโภคจากไมโครบล็อกอย่างอัตโนมัติ โดยระบบประกอบไปด้วย 5 ส่วนหลัก ได้แก่ (1) ส่วนเก็บรวบรวมไมโครบล็อก (2) ส่วนกลั่นกรองความคิดเห็นจากข้อความทั้งหมด (3) ส่วนตรวจสอบขั้วความคิดเห็นของแต่ละข้อความ (4) ส่วนแบ่งหมวดหมู่ผลิตภัณฑ์ และ (5) ส่วนสรุปผลและนำเสนอผลลัพธ์ทั้งหมด

กระบวนการวิจัยประกอบไปด้วยการวิจัยเชิงปริมาณและการวิจัยเชิงคุณภาพ โดยเริ่มต้นจากการจัดสนทนากลุ่มเพื่อหาความต้องการเชิงธุรกิจ และการสัมภาษณ์เชิงลึก 5 ครั้งเพื่อหาทิศทางในการพัฒนาซอฟต์แวร์ จากนั้นนักวิจัยได้เลือกเพียงบางหน้าที่ที่สำคัญเพื่อพัฒนาต้นแบบโดยใช้ความคิดเห็นเกี่ยวกับสมาร์ทโฟนเป็นกรณีศึกษา และเก็บรวบรวมข้อมูลจากทวิตเตอร์เพราะเป็นตัวแทนที่เหมาะสมของไมโครบล็อก สุดท้ายเพื่อตรวจสอบผลของต้นแบบนักวิจัยได้จัดสนทนากลุ่มขึ้นโดยประยุกต์ใช้โมเดลการยอมรับเทคโนโลยี ได้ผลการรับรู้ประโยชน์และผลการรับรู้ความง่ายในการใช้งานเป็น “เห็นด้วย” และทัศนคติการใช้งานเป็น “เห็นด้วยอย่างยิ่ง”

ผลลัพธ์การศึกษาครั้งนี้คือต้นแบบของเครื่องมือการตลาดเชิงนวัตกรรมซึ่งประยุกต์หลักการเหมืองความคิดเห็นบนไมโครบล็อกเข้ากับหลักการของ Johari Window ในการตอบโจทย์ความเข้าใจผู้บริโภค เพื่อเป็นแนวทางสำหรับคนในอุตสาหกรรมซอฟต์แวร์ในการข้ามขอบเขตมายังการทำวิจัยเชิงการตลาดต่อไปในอนาคต

สาขาวิชา ธุรกิจเทคโนโลยีและการจัดการนวัตกรรม ลายมือชื่อนิสิต .....

ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก .....

ปีการศึกษา 2554 ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์ร่วม .....

##5087841220 : MAJOR TECHNOPRENEURSHIP AND INNOVATION MANAGEMENT
KEYWORDS: OPINNION MINING / SENTIMENT ANALYSIS / SOCIAL MEDIA / TWITTER / MICRO-BLOG

WILAS CHAMLERTWAT : INNOVATIVE MARKETING TOOL BY APPLYING OPINION MINING ON THE MICRO-BLOG. ADVISOR : ASST.PROF. PATTARASINEE BHATTARAKOSOL Ph.D., CO-ADVISOR : TIPPAKORN RUNGKASIRI, Ph.D., 114 pp.

Traditional marketing research approaches may not adequate for today business competitions because they require a large amount of time and resources. As a supportive solution, the researcher proposes a system called the Micro-blog Sentiment Analysis System (MSAS), based on the opinion mining concept to automatically analyze consumer opinions from the micro-blog. The MSAS consists of five main modules to (1) collect micro-blog posts, (2) filter opinion posts from entire posts, (3) detect polarity in each post, (4) categorize product features, and (5) summarize and visualize the overall results.

The research process composed of both qualitative and quantitative approaches; starting this study from the focus group discussion to scope the business requirements. Then, five in-depth interviews were conducted to find out the direction of the software development; researchers had chosen some important functions to implement the software prototype by using the smartphone as the case study and collected posts from Twitter as a proper representative of the micro-blog service. Lastly, the research process did focus group discussion to verify the result of the software prototype by adopting Technology Acceptance Model in this session. The final result of both Perceived Usefulness and Perceived Ease of use is “agree” while the attitude toward using this software is “strongly agree”.

The outcome of this study is a prototype of innovative marketing tool that applying the opinion mining concept on the micro-blog with Johari Window to answer the consumer insight question. It will be a guideline for people in the software industry to cross the disciplinary to the marketing research field in the future.

Field of Study : Technopreneurship and Innovation Management

Academic Year : 2011

Student’s Signature …….

Advisor’s Signature …….

Co-advisor’s Signature ……...

ACKNOWLEDGEMENTS

I dedicate this study to my parents, Pipat and Pajongjit Chamlertwat, who have made my education to this point of success.

Since this study consists of both qualitative and quantitative research, there are so many lovely people involved in each process. First, I am heartily thankful to Assistant Professor Pattarasinee Bhattarakosol, Ph.D., Tippakorn Rungkasiri, Ph.D., and Choochart Haruechaiyasak, Ph.D. for their warm guidance and support. Their encouragement and accurate comments were critical importance for this work.

Second, I would like to thank you all of my developers for their help in coding the prototype. They are Warut Surapat, Phattarakorn Singsopha, Kanokporn Nonsrichai, Kornthip Lim, and Kornthip Nakchamrassri.

Next, I sincerely appreciate information and knowledge from all interviewees and participants in both focus group discussions. They are Pongskorn Saipetch, Ph.D., Kittipong Kittipassorn, Pong Sakuntanak, Pawoot Pongvitayapanu, Prasert Sriuranpong, Chanin Visarnbuchanee, Nattakit Tangpoonsinthana, Ph.D., Witawat Jayapani, Nathida Ratthanawut, Kasidit Kolasastraseni, Oranuch Lerdsuwankij, Pim Soonsawad, Ph.D., Chatchawit Aporntewan, Ph.D., Poomjai Krittiyanon, Savisith Phuawong, Teerawat Suwanpinit, Jiraz Pipatwasin, Worawisut Pinyoyang, Phatthaworn Phongphaew, and Pnern Asavavipas.

Also, special thanks to Parkpoom Tripatana and Pensri Arunwatanamongkol to let me prioritize this study rather than my responsibilities in the companies.

Lastly, I offer my regards to all of those who supported me in any respect during the completion of the dissertation included those who are not mentioned above. Million thanks for you all.

CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

Terminology	Definition
Application Programming Interface (API)	A set of programming instructions and standards that a software program implements in order to allow other software to interact with it
Information Gain (IG)	A common feature selection technique using in text classification
Information Retrieval (IR)	The science of searching information to reduce what has been called “information overload”
JavaScript Object Notation (JSON)	A lightweight data-interchange format used primarily to transmit data between a server and web application, serving as an alternative to XML
Natural Language Processing (NLP)	A field of study concerned with the interactions between computers and human (natural) languages
Really Simple Syndication (RSS)	The standardized web feed syndication that used to publish frequently updated works such as blog entries, news headlines, and video
Support Vector Machine (SVM)	A set of related supervised learning methods used for classification and regression to predict whether a new example falls into one category or the other
User-Generated Content (UGC) or Consumer Generated Media (CGM)	A term used to describe any form of content that is created by consumers or end-users of an online system or service and is publically available to others

Uniform Resource Locator (URL)	The global address of documents and other resources on the World Wide Web
eXtensible Markup Language (XML)	A markup language that defines a set of rules for encoding documents in a format that is both human-readable and machine-readable

CHAPTER I
INTRODUCTION

Rationale of the Study

Currently, innovation is “the must” to compete in high rat-race business world. For marketers, understanding in “consumer insight” or “voice of consumer” is one of the top challenging tasks. It is not straight forward to find out what people think, when they want to adopt new products, which features is desired or undesired, and how people feel about it. Moreover, the evaluation time and cost are important factors for this kind of consumer research. With traditional approaches, which may obtain information from surveys, interviews, or observations, a long time interval on the research process is unavoidable while a high expense in the evaluation is the consequence. Therefore, results from the traditional marketing research will not be able to effectively respond to the high business competition situation any longer.

Businesses need a novel approach to learn their customers. Since consumer behavior is always changed, businesses want to gain the actual on time feedback regarding key success factors, or failure factors from consumers to adjust their strategy. Especially for any short-life cycle industries, such as Consumer Electronic (CE) devices, time is very crucial circumstance because there is very short period to win the market share before the product is obsolete. Every business requires a window of opportunities to penetrate into consumer’s insight. As a result, an innovative marketing tool to understand and to serve consumer needs is very essential.

The social media is now becoming a significant alternative channel for communicating, sharing and networking, both for interpersonal and business purposes. It has abruptly diffused to people daily life. The Social Networking Service is not only a tool for business to speak out, but also a rich source for businesses to watch their customer movement. Although people avoid answering the fact of what they have thought to business representatives, they incline to express those things in the Internet to their friends through social networking sites. When people post about products, they

are willing to share their actual opinions and sincere comments about what they love or hate. Especially, communication on the micro-blog is interesting resource. Many people use micro-blog on daily basis to talk about whatever topics happened in their lives. Messages in the micro-blog are in unique pattern that is either concise or noisy. It will be very useful and challenging for the company to transform the information from this source to explicit knowledge for business improvement.

Opinion mining (OM) is the study related text that contains opinions, emotions or sentiments. It is an emerging research area which combines between computational linguistics and Information Retrieval (IR). OM can lead to automatic determine, analyze and generalize of consumer insight. The challenge is how to detect the useful opinions that any businesses want to know using the Natural Language Processing (NLP) technique. Unlike most previous researches, which retrieve raw data based on blog, forum, or reviewing website for textual processing of factual information, this study focuses on the opinion mining on the micro-blog. From our literature, there is no research applied OM on the micro-blog concept to directly assist the marketers in the company to understand their positioning in consumer mind, and to adjust their products/services to meet the consumers' needs.

To address this research gap, the research implemented software prototype called "Micro-blog Sentiment Analysis System (MSAS)" for handling with opinion mining on the micro-blog for a specific domain, which smart phone industry was chosen as a case study.

Research objectives

The thesis statement is "Understanding consumer opinion through the micro-blog analysis can turn into business services." Therefore, this study has three research objectives:

1) To design and develop a marketing tool by applying opinion mining concept on the micro-blog for business purposes

2) To implement software prototype corresponding to the above concept

3) To study and propose the business feasibility for above software

Scope of research

In this study, the researcher would like to associate sentiment expressions with the product features that retrieved from the micro-blog. However, the result of this study is not for general uses. To adapt for other purposes, researchers have to prepare a new corpus.

Below is the scope to declare in this study:

1) The design of software prototype is suitable for decision support only, not for decision making since there are many other factors influence consumers' behavior.
2) This model works well with English language only.
3) The studied product must have been announced in the Internet in advanced; therefore, software can collect conversation from the micro-blog.
4) The reliability depends on number of analyzed messages, which is relied on the collecting time and the limitation of the micro-blog service permission.

Limitation of research

While this study may provide useful information on the opinion mining for the smartphone industry, there are some limitations which need to be addressed.

1) Language: Although this study was conducted in Thailand, the researcher decided to collect the micro-blog data in English because of lacking of micro-blog messages in Thai at the time the research was started. In 2010, Thai people just started using Twitter. So, there were not enough data for the experiment. Albeit, the focus group discussions and in-depth interviews were conducted with Thai experts. It may limit the generalizability of the results.
2) Computation capability: The researcher conducted the experiment using WEKA software on a PC which took too long time to process. Therefore, the researcher decided to select sampling from our collected data to represent the whole population.

Despite the limitations of the study, it is believed that this research offers

important implications for Thai opinion mining research and practice by providing information on the smartphone industry as a case study.

Terminology

Table 1.1 Definitions

Term	Definition
Consumer Insight	Information about unidentified or unmet need of consumers in the market.
Corpus	A large and structured selected collection of texts using for a linguistic analysis.
Micro-blog	One of social networking services that provides short note facility to update user status.
Opinion Mining (or Sentiment Analysis)	The text analysis process to determine the attitude, emotion, or affect in a document.
Social Networking Service	An Internet-based application that provides tool for users to interact with other people.
Synset	A group of rational synonym that is considered semantically corresponding.

Contribution

Since Technopreneurship and Innovation management Program requires accomplishing both academic and business goal, therefore, the followings are our main contributions of this work:

1) For academic side, the researcher proposed a cross-disciplinary model that influences in the real-world, large-scale scenarios using a set of micro-blog data as our case study.

2) For application practice, the researcher developed a system for a business to detect and extract opinions from consumers in order to identify the success or failure factors for products or services.

Methodology

As illustrated in Figure 1.1, the researcher separated the task into four parts as follows.

1) Focus group discussion - to identify the need of a new marketing tool on the micro-blog
2) In-depth interview - to find the direction of software prototype development
3) Software experiment - to implement software prototype for the experimental study
4) Focus group discussion - to verify the result of software prototype and the innovation acceptance

Structure of the dissertation

The organization of this dissertation document is that Chapter I gives an introduction to the topics. Chapter II deals with the theoretical foundations relevant to the problem areas. This chapter reviews Micro-blog, Opinion Mining, and Business Intelligence as they are the central issues to the development of this study. Chapter III provides an overview of the complete methodology. Chapter IV presents the implementation and results of the study. Chapter V describes business feasibility study for the software commercialization. Finally, Chapter VI summarizes the work and presents some suggestion for the future work.

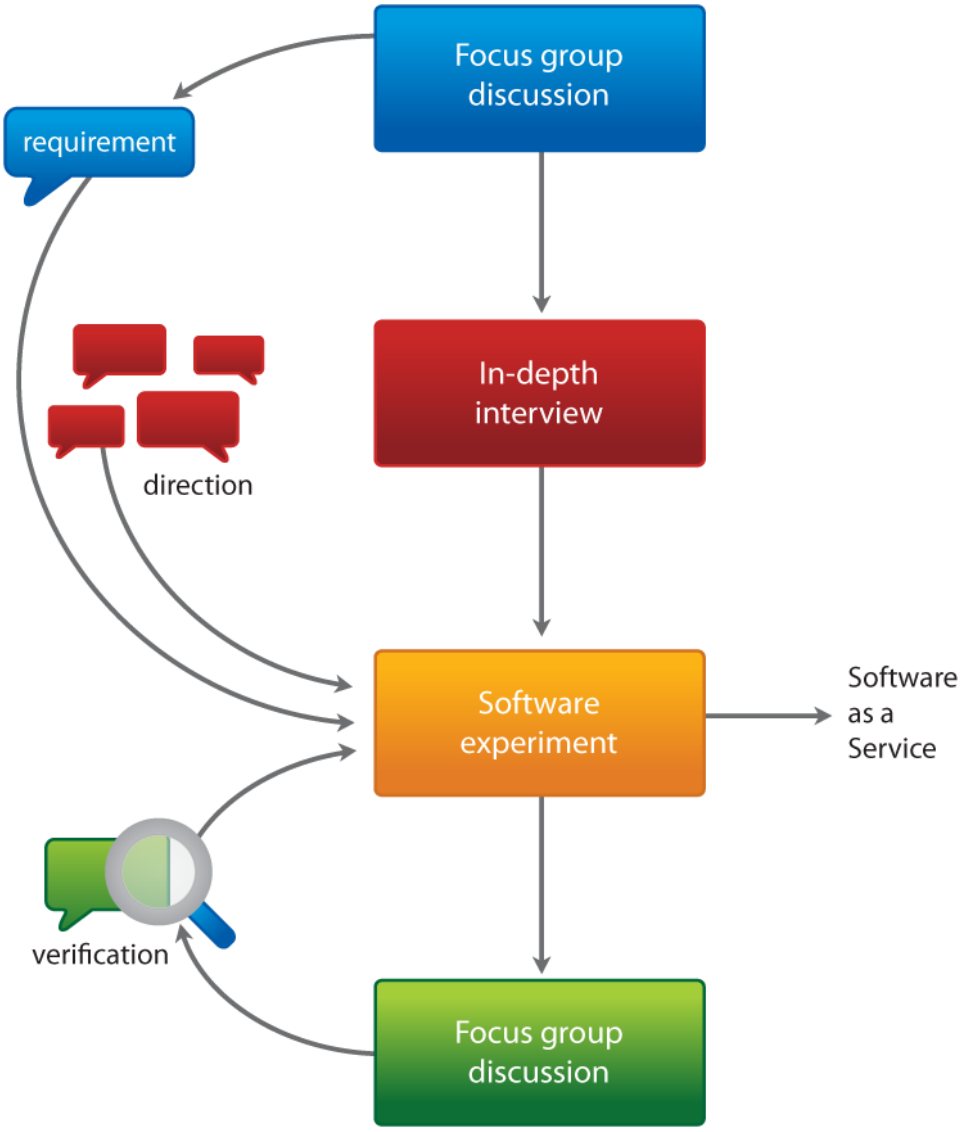

Figure 1.1 Research methodology

CHAPTER II
LITERATURE REVIEW

This chapter presents background information which is fundamental for understanding the work accomplished in this study. It is separated into three main parts, which are the Micro-blog, the Opinion Mining, and the Business Intelligence.

Micro-blog

The micro-blog is always described as a smaller version of the weblog, usually recognized as a part of the “Web 2.0”. Since the micro-blog enriched with features for social networking and with a strong focus on mobility, it has emerged and grown with an amazing rate. The micro-blog system first is aware from the tipping point of Twitter in 2006 and similar systems, including Jaiku, Pownce, and others.

A. Overview of the micro-blog

The micro-blog provides an informal communication channel that allows people to express and share their information about present situations, considerations, or estimations either to public or just in private group. Its main objective is to allow users post any update messages with small effort. This kind of services often supports a variety of usages. After registering, users will gain their own public space where they can post any short messages. Then, other users will be able to subscribe, or so called “follow”, by adding to the “following” list. Similar to blog, the up-to-date messages of friends will be showed in chronological order on each user home page. Moreover, these messages will be spread via mobile phones, email or web.

The micro-blog diverges from conventional blog in both how and why people use it. Compared to a regular blog, the micro-blog can fulfill a need of a faster and more instant manner of communication. Although the micro-blog limits number of characters in each message to fit in a single Short Message Service (SMS), it benefits users by reducing time and thought to create contents. This also brings about the effortless and

flexible to type whatever users think during their daily life on mobile phones. Its easiness boosts more frequent postings. A productive user may update his or her blog every few days, whereas he or she might post on the micro-blog every few hours or minutes.

The micro-blog users can update their statuses, express their own ideas, tell other people about problems they are facing, and publish solutions or some interesting articles. In addition, the researcher found that the micro-blogging is applied to some uses, such as the weather report service, and the traffic report service. Although some contents for example “I’m having fried rice for this dinner” might sound unimportant to public, this message may be valued for some friends or family members. With the current reputation of the micro-blog usage, marketers should not ignore to understand how and why people use it. The understanding of the micro-blog practice will help researcher developing practical software to serve needs of marketers and people who involved.

B. Micro-blog characteristics

Most of micro-blog applications share common characteristics (Zhao and Rosson, 2009) as described below:

1. Limited-length message.

   This is a very important characteristic of the micro-blog. With this property, it promotes the user capability to update their statuses with minimal effort, which quite different from the blog that often requires the user to spend more time on it. As a result, the micro-blog users tend to update their statuses much more often.

2. Various input and output channels.

   Most micro-blog services provide a variety of input and output tools, and lead them to be available over various devices. For example, the message can be posted directly from the web site. The users can update their statuses via an Internet-connectible mobile device. For obtaining news, users who subscribe to receive the update from their friends can view the update on the web, or receive the update via the Really Simple Syndication (RSS). Moreover, the developers

are provided with the Application Programming Interface (API) to obtain the capability to use the micro-blog service in their software.

3. Wide range of information in many domains.

Due to the large number of users and its diversities, contents of messages cover many domains. This is a very rich resource for the research study.

4. Broadcasting manner.

The micro-blog can be viewed as another type of SMS. However, the main difference is that the micro-blog is published in the broadcasting manner. By default, messages can be accessed by everybody unless users decided to protect their update statuses.

5. An immense number of noises.

There are some reasons that bring the micro-blog a high portion of noise. For example, it covers the large number of domains; users often update their statuses. As a result, there are a bulk number of messages over the application. Furthermore, the micro- blog does not prohibit the act of advertising and mining. On the micro-blog application, one can obtain the update from others after he subscribed to the people of interest. The subscription is the bi-direction relation.

C. Twitter

There are many online applications that were implemented with the micro-blog concept. Twitter, launched in July 2006 by Jack Dorsey, has enormous possibility to become a collective source of intelligence that can be used to uncover people ideas, opinions, and sentiments. It asks users with an easy question that “what’s happening”. Then, users update their statuses (known as a “tweet”) to spread across the globe immediately. Twitter as the most popular micro-blog service has gained reputation worldwide and sometimes called the “SMS of the Internet” (D'Monte, 2009).

The tipping point of Twitter popularity was the South by Southwest (SXSW) festival in March 2007. Twitter usage went up triple from 20,000 tweets per day to 60,000 tweets per day immediately. Response at the festival was awesomely optimistic.

Scott Beale, Laughing Squid blogger, said that Twitter "absolutely ruled" SXSW while Danah Boyd, the social software researcher, said that Twitter "owned" the festival (Terdiman, 2007). As seen in Figure 2.1, Twitter has rapidly grown both the number of tweets and number of users from 2007 to 2010.

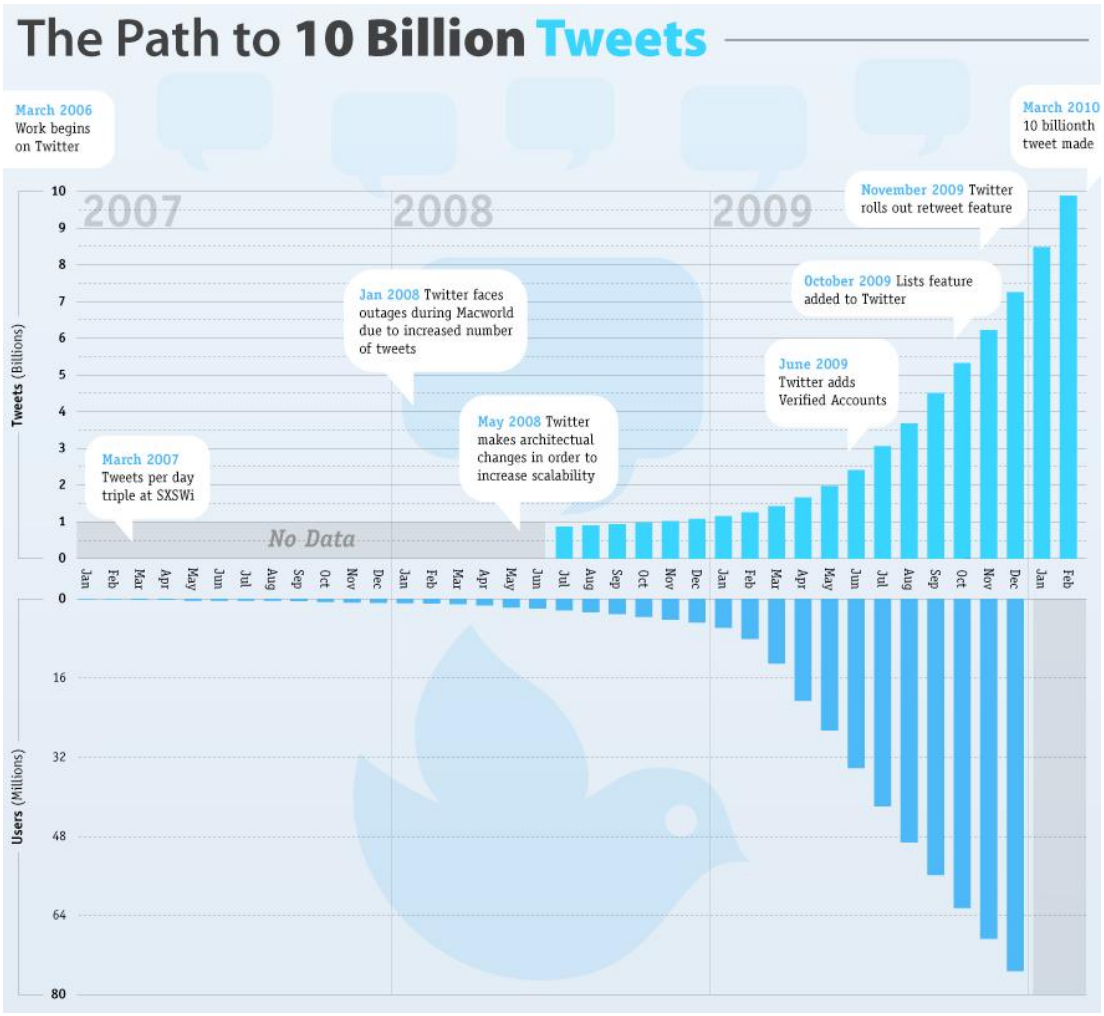

Figure 2.1 the growth of Twitter from 2007 to 2010

In November 2008, Facebook, the most popular Social Networking Service, offered $500 million in shares for Twitter. This is a remarkable amount for a service with estimated 6 million users worldwide at that time, whilst in the middle of a recession (Kazeniac, 2009). It seems that Facebook sees rising importance in the subject and has high expectations for its role in the future web. However, at the end the deal was not happened. In March 2012, Twitter celebrated 6 years anniversary with more than 140

million active users from 500 registered accounts. They have recorded 340 million tweets per day.

1. Tweet

Tweet is a term for a message posted on Twitter. In general, a tweet likes a concise blog post. Its limitation of 140 characters was set for SMS compatibility. This maybe a reason that bring about Twitter users to post quite often, and brings a culture of abbreviation and slang frequently abused in SMS to the Twitter. Therefore, the enormous shorthand writings which are unpredictably steady across users in Twitter would be seen very often.

Moreover, tweets often contain a Uniform Resource Locator (URL) which leads the reader to outside resource, such as article, blog post, or other websites. The URL shortens service like bit.ly, goo.gl, tr.im, and others are used in order to help putting the URL under the limited length feasible. For example, http://bit.ly/17zKu is abridged from http://en.wikipedia.org/wiki/Microblogging. From March 2010, Twitter has provided t.co, link shortening service, for links posted in Twitter to track the number of clicks on that link and also to protect users from malicious websites.

The public appearance is the default tweet setting of Twitter. This permission enables every user to subscribe or follow other users and read each other messages without the need of providing 2-way consent. Each Twitter user has a profile page, which all their tweets are collected into one chronological order. Besides showing on user profile page, tweets will be spread to subscribers or followers list via SMS, instant messaging, email, and other applications, such as Hootsuite, Tweetdeck, and UberSocial.

As a broadcast medium, Twitter delivers tweet to every user in the network. However, Twitter also provides special syntax to send tweet directly to a specific user, or use it as a reply to response to an earlier mention. By typing a “@” symbol to the intended username such as “@LadyGaga”, the user LadyGaga will receive this message in her Reply section. However, the @

symbol is opened for public; everybody will be able to read this message. If user would like to send a private message to other users, they can perform by adding DM followed by @username such as DM @LadyGaga.

Twitter has an interesting feature to spread the message in the system. This is called “Retweet” or “RT”, which implies to the redistribute action of the message by another user. For example, Alice posts “Happy new year”. To repost another user message, Bob can work in two ways. First Bob can retweet by type “RT @Alice: Happy new year”. Second, new Twitter retweet is much easier by clicking the retweet link in the web and the tweet will, then, be forwarded to all of user followers.

One more interesting symbol is # or so called “hashtag”, which is used to freely categorize tweets together. Users can label using of hashtag in front of word. For example, “Can’t wait the new #iPad”. The practice of using hashtags may stem from a HTML anchor point or prefacing specialized words with punctuation marks in a computer programming (Boyd *et al*., 2010). Apart from these issues, tweets in Twitter are full of misspelling and wrong linguistic sentences, which make it quite difficult to study.

Last but not least, Twitter also provides Application Programming Interface (API) to allow software developers to retrieve any information from Twitter. This API service is available in both Extensible Markup Language (XML) and JavaScript Object Notation (JSON) formats. Users will be able to tweet from third party clients; desktop applications, or mobile application. Besides, users will be able to monitor or search tweets from the Twitter search API by keywords, hashtags, or usernames, which is very important for this study.

2. Follow

The relationship in Twitter is called “follow.” Users are allowed to follow any other users in the system. Once they follow a user, they will receive tweets posted by that user. However, each user can protect their privacy by setting permission to mutual relation only. Users say that they are “following” someone if

they subscribed to him. Also, users can say that they have “followers” if there is someone who decides to follow them. Supposed there are two users; Alice and Bob. If Alice followed to Bob, she would see all updates from Bob. Bob, conversely, will not see any updates from Alice unless Bob follows her back. In this case, Alice is following Bob. For Bob, Alice is follower.

Table 2.1 Top 5 most popular Twitter users by number of followers

Profile image	Name	Number of followers	Number of followings	Number of tweets
	Lady Gaga @ladygaga	22,634,087	139,273	1,367
	Justin Bieber @justinbieber	19,957,179	123,098	14,975
	Katy Perry @katyperry	17,624,861	85	3,924
	Rihanna @rihanna	16,780,932	777	4,812
	Shakira @shakira	15,420,468	59	1,310

As shown in Table 2.1, here is the list of Top 5 most popular Twitter users by number of followers in April 2012. It can be noticed that one of the major reasons of Twitter popularity may come from this asymmetrical relationship among users. It creates an interesting social network structure to form connections of followers and friends. Moreover, Twitter has a basic rule to limit the number of 2,000 maximum following in total. The number of additional users is different for every user and is based on the ratio of followers to following. On the contrary, Twitter does not limit the number of people who can follow others. Therefore, each user can have as many as followers, but is controlled the number of following depended on a relationship ratio. This constraint is set to prevent spamming and improve the system performance.

3. User interface

Twitter changes its user interface many times; the latest one was happened in November 2011. As shown in Figure 2.2, it is an example of a user interface from http://twitter.com/vajiramedhi before the user logins. Figure 2.3 shows the user interface after the user login. This latest design aims to provide Twitter's user experience (UX) consistent across all platforms, from a mobile to a desktop. On the top navigation tab, there are three main menus titled Home, @Connect, and #Discover. Home is where the user can view tweets from other people that he/she has chosen to follow. @Connect section is the place to see all recent activities surrounding the user, such as who has followed, mentioned, retweeted or favorited the user. #Discover offers users an expedited way to consume the latest information in a glance, reflecting the user interests based on the current location.

Figure 2.2 Twitter.com user screen before login

Figure 2.3 Twitter.com user screen after login

4. State-of-the-art research

(Java *et al.*, 2007) pointed that a user's intention over Twitter can be classified into four categories according to the link structure as following.

a) Daily chatter - users tweet to update their daily activities; discuss events in their lives or their current thoughts

b) Conversations - users tweet as a tool for conversation by direct the

message to target using @ mark.

c) Sharing information - users tweet to collect and post the link they are interested in.

d) Reporting news - users tweet to report the news, including commenting on current events

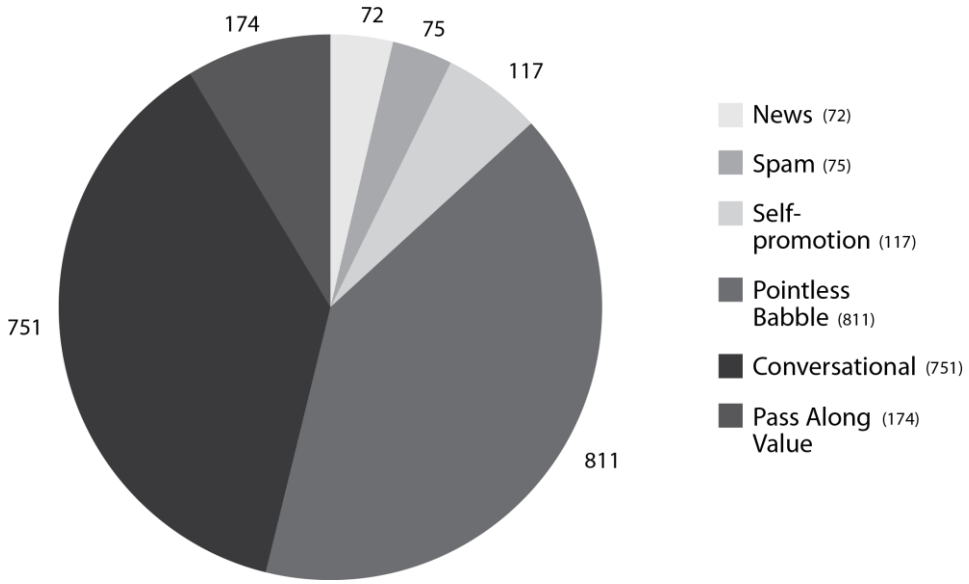

Figure 2.4 Content of tweets

As shown in Figure 2.4, the topics of tweets are varied from something happened in the daily life to news stories, and other issues. (Pear Analytics, 2009) studied 2,000 English tweets retrieved from the U.S. over 2 weeks. These tweets can be classified in six categories as follows.

- Pointless babble - 41%
- Conversational - 38%
- Pass-along value - 9%
- Self-promotion - 6%
- Spam - 4%
- News - 4%

In response to this result, Danah Boyd argued that the Pear Analytics labeled "pointless babble" should be better categorized as "peripheral

awareness" or "social grooming" since some messages may contain interesting contents to which the researcher may not be familiar.

(Java *et al.*, 2007) also identified three main categories of Twitter users as following:

a) An information source is a group of users who updates often and has various people to whom are subscribing. However the number of the people whom this user will subscribe back is less. This category includes both real human and the automated tools.

b) Friends are the common users who use Twitter to keep listening to their friends' activities.

c) An information seeker is a user who rarely updates his or her status but often listens to others.

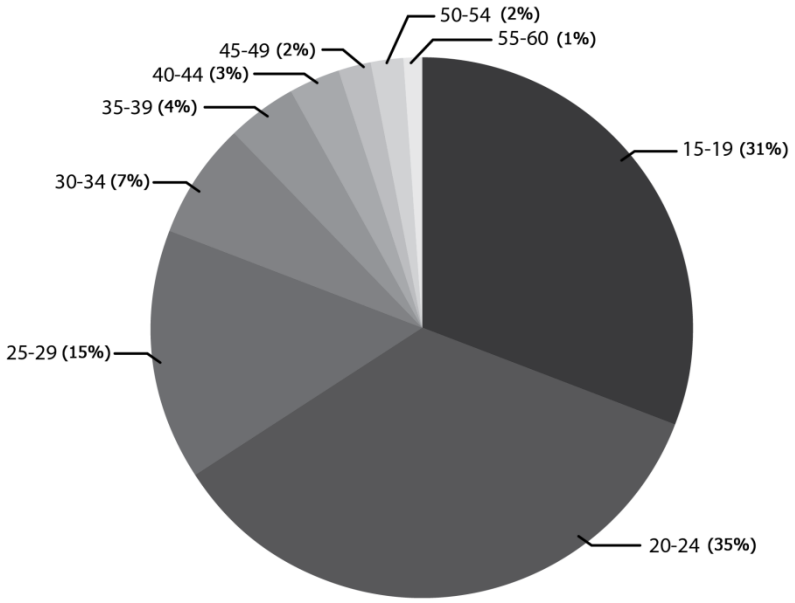

Figure 2.5 Twitter users categorized by the aging group (years old)

As illustrated in Figure 2.5, (Cheng *et al.*, 2009) studied the ages of Twitter users based on a sample set of users who revealed their ages and found that 65% of Twitter users are under 25 years old. Moreover, the group of young people increases continuously. As Twitter users attract more followers, they tend to tweet more often. This is particularly evident that the average number of

tweets per day will rise from three to six when the user has reached 1,000 followers. Also, if the user reaches the number of followers more than 1,750, the number of tweets per day will soar to ten.

Country	% twitter user
USA	50.88
Brazil	8.79
UK	7.20
Canada	4.35
Germany	2.49
Indonesia	2.41
Australia	2.39
Netherlands	1.32
India	1.27
Japan	1.22
Mexico	1.11
Philippines	1.08
France	0.98
Spain	0.78
Singapore	0.69
Italy	0.65
Ireland	0.52
Chile	0.51
Sweden	0.50
New Zealand	0.47

Source: sysomos.com sysomos

Figure 2.6 Percentage of Twitter users by country

As illustrated in Figure 2.6, the research from Sysomos in 2009 stated that the U.S. has the greatest share of Twitter users with 50.88% followed by Brazil with 8.79%. The gap between the first and the second is quite abundant. UK was ranked the third with 7.20%. One interesting country is India, which has the percentage of Twitter users at 1.27% more than Japan at 1.22%. By the way, the percentage of Twitter users of Thailand is not even 1%.

Opinion Mining

The Opinion Mining (OM), or so-called Sentiment analysis, is basically the approach of detecting feelings of the writer regarding some specific topic. As a field of research, OM is closely associated with computational linguistics, Natural Language Processing (NLP), and text mining that analyzed not only what the topic is about, but also with the opinion it expresses. Proceedings from the study of affective state (psychology) and judgment (appraisal theory), this field seeks to answer questions long studied in other areas of discourse using new tools provided by the data mining and computational linguistics. Functional to the extraction of opinions from text is the determination of the relevant entities of the language that are used to express opinions, and their opinion-related properties. For example, determining that the term beautiful casts a positive connotation to its subject.

A. The importance of Opinion Mining

Opinions or sentiments have a huge impact on human's walk of life. People use opinions for both expressing their points of view and listening from other people about how they perceive something. People are accustomed to ask and compare opinions from others to support their own decisions, such as recommending food, requesting job references, or consulting for the mobile phone purchase.

Organizations and industries are also benefit from opinions. When an organization needs finding opinions of the general public about its products or services, marketers will conduct surveys or the focus group discussion. For example, a company may perform an annual market surveys to collect people point of views about the products or services, and also information about their competitors in order to compare and plan their marketing strategies.

Presently, there are many sources, from which opinions can be gathered, such as questionnaire, interview feedback, and the Internet. With the Internet, especially with the explosive growth of the User-Generated Content (UGC) on Webs, the world has

changed. It is probably the most valuable one, given its widespread diffusion, accessibility, and liberality. Everybody can post reviews of products/services at corporate websites and express views on almost anything in personal blogs, forums, and social networking sites. More and more people are making their opinions available to strangers that are neither their personal acquaintances nor well-known professional critics via the Internet.

For an organization, researching on the Internet is an alternative choice beside to conduct surveys, to organize focused groups or to employ external consultants in order to find consumer insight about its products/services and those of their competitors. This practical need for opinion analysis tools has met the interest of researchers working in the fields of Information Retrieval and Computation Linguistics. (Hu and Liu, 2004) said that it is very difficult for a human reader to find relevant sources, extract sentences, summarize and organize them into usable forms. An automatic processing tool for the opinion mining and the summarization system is thus required.

B. Opinion Mining background

Currently, OM is not the only term used in literatures; many other names have been referred to such as “Opinion Analysis”, “Subjectivity Analysis”, and “Sentiment Analysis”. All these names can be considered more or less interchangeably as synonyms to generally identify the discipline. Moreover, each name would possibly denote some specific subtasks (Pang and Lee, 2008).

OM is concerned with the analysis of opinions expressed in documents. OM is, thus, a non-topical text analysis discipline, which is not concerned with the topic of the analyzed documents but with the opinions expressed about its subject matter.

Since 2001, it seems to mark the beginning of widespread awareness of the research problems and opportunities in OM. Therefore, there have been several of papers published on this discipline. Factors behind this awareness may be included the rise of Machine Learning methods in NLP, the availability of datasets for machine learning algorithms to be trained on, and the development of review aggregation

websites on the Internet.

(Dave *et al*., 2003) said that the ideal opinion mining tool would “process a set of search results for a given item, generating a list of product attributes (quality features, etc.) and aggregating opinions about each of them (poor, mixed, good)”. Much of the subsequent research self-identified as opinion mining fits this description in its emphasis on extracting and analyzing judgments on various aspects of given items. However, (Liu, 2011) has interpreted the term more broadly to include many different types of analysis of evaluative text.

Recent works in OM have begun to apply fine-grained semantic distinctions focused mostly on interpretation of metaphor, narrative, point of view, affect, evident in text, and related area. Such methods, however, require the construction of large and complex lexicons, giving values for multiple sentiment-related attributes to many different lexical items. For example, a key attribute is the type of attitudes is expressed by a lexical item, e.g. beautiful expresses appreciation of an object’s quality, while evil expresses a negative judgment of social behavior.

Unlike in usual topical analysis, the opinion statement authorship can be integral to the problem. One of the main problems is quotation. It is important to know that the opinion expressed in the document is representative of the actual intent of the author. Political commentaries and news are full of quotations and opinion citations, and can have a complex structure that is difficult to be recognized. For example, a news article about a political debate would have a mix of quotations from the debaters, the pundits commenting on the debate, and perhaps even the author’s stance on the issues.

As mentioned before, some of the most studied texts in OM are product and movie reviews (Popescu and Etzioni, 2005). (Pang and Lee, 2004) proposed a machine-learning method that applies text-categorization techniques to examine the relation between a subjectivity detection and a polarity classification, showing that the subjectivity detection can compress reviews into much, shorter extracts that still retain polarity information at a level comparable to that of the full review. Meanwhile, one of the most difficult areas for OM methods is that of politics. (Gamon *et al*., 2008) claimed that

political discussions are fraught with quotations, sarcasm, and complex references to persons, organizations, and ideas. Some works have been performed on determining whether a political speech is in the supporting of, or in the opposition to, the issue under debates (Bansal *et al*., 2008; Thomas and Pang, 2006). There is a related work on categorizing election forums into “likely to win” and “unlikely to win” (Kim and Hovy, 2007).

C. Text Mining

Like the data mining, the text mining is the process of analyzing patterns of unstructured text instead of a structured database. The data mining is the extraction of implicit and possibly useful information from a database or a data warehouse. With the text mining, conversely, the information to be extracted is explicitly and clearly stated in the text, but it is difficult to an automatic process by a machine (Witten, 2005).

1. Preprocessing

For this study, the researcher defined preprocessing as all tasks that take place before main functions.

a) Tokenization is a method to segment a sentence into words by considering from the word delimiters, such as space, and punctuation marks. For example, a sentence "Just realised that the screen on the iPhone is the same size as my android phone" is tokenized to 16 words, "Just", "realised", "that", "the", "screen", "on", "the", "iPhone", "is", "the", "same", "size", "as", "my", "android", "phone". All tokenized words form a bag of words.

b) A regular expression is a pattern used to process a certain text for almost any programming languages, such as string extracting, character modifying, and word comparing. For example, zipcode from "Bangkok 10400" can be extracted with this following pattern /([0-9]{5})/ .

c) Stemming is a very useful mechanism in the field of Information Retrieval for reducing English words to their root form. For example,

the word “consider” is stemmed from “considering” or “consideration”.

2. Feature selection

Feature selection is one of the most significant tasks in the text mining in order to choose the relevant data for building robust learning models in the training set. The best subset contains the least number of dimensions that most contribute to accuracy. Selecting the terms has many advantages. For example, a reduced term size speeds up the training process. It also increases the classification accuracy because of removing the irrelevant features.

a) Bag of Words (BOW)

It is a set of words that contains every words used in the document, with multiple appearances of a word showing multiple times. It is a simplifying assumption used in NLP. In this model, a text is represented as an unordered collection of words, disregarding grammar and even word order.

b) Information Gain (IG)

It is selecting features that reveal most information about classes. It is a measurement of the commonality of a feature in a particular class compared to the commonality of it in all other classes. A word that occurs primarily in positive posts and rarely in negative posts is high information. For example, the presence of the word "magnificent" in a movie review is a strong indicator that the review is positive. That makes "magnificent" a high information word.

c) Chi-square (CHI)

Chi-square, or sometime referred to X^2, is another popular feature selection method. It tests a null hypothesis stating that the frequency distribution of certain events observed in a sample is consistent with a particular theoretical distribution. The events considered must be mutually exclusive and have total probability.

3. Classification methods

Any classification method uses a set of features to characterize each object, where these features should be relevant to the task at hand. The researcher considered these methods for supervised classification, meaning that a human expert both has determined into what classes of an object may be categorized and also has provided a set of sample objects with known classes. This set of known objects is called the training set because it is used by the classification programs to learn how to classify objects. There are two phases to constructing a classifier. In the training phase, the training set is used to decide how the parameters ought to be weighted and combined in order to separate the various classes of objects. In the application phase, the weights determined in the training set are applied to a set of objects that do not have known classes in order to determine what their classes are likely to be.

a) Naïve Bayes (NB) is a simple probabilistic classifier based on applying Bayesian theory by the normal distribution to model numeric attributes. It is particularly suited when the dimensionality of the inputs is high.

b) Support Vector Machines (SVM) has been emphasized as one of the best performance. It is the method that analyzes data and recognizes patterns, used for classification and regression analysis. The essence of SVM is to find a hyperplane that separates document vectors from one class to the other as much as possible.

D. Sentiment Analysis

As mentioned earlier, the sentiment analysis can be swop with the term OM. The sentiment analysis turns out of the need for an automated opinion disclosure and a summarization system dealing with a large amount of data to allow a machine to understand human generated content (Lake, 2011). In business arena, this technology has been used in mining opinions for brand monitoring, polls, financial trading, marketing and many more in real world problems.

The Subjectivity classification and the Sentiment classification are perhaps the most widely studied topics in this field. The Subjectivity classification is a process to separate subjective from objective sentences or distinguish opinions from facts while the Sentiment classification is a process to determine sentiment orientation whether that sentence expressed positive or negative feeling. Also, some researchers are interested in determining the intensity (strength) of sentiment polarity to measure the semantic intensity. Feature-based sentiment analysis is an in-depth study about the determining of the expressed sentiments on different features. For example, the feature-based sentiment analysis of the smart phone screen is the study about people expression on the screen whether it is positive or negative.

Various sentiment analysis methods have been found out in several different levels of text granularity (Agarwal *et al.*, 2011). The coarse-grain level starts from a document level classification task (Turney, 2002) to a finer-grain level of a sentence (Hu and Liu, 2004) and at the phrase level (Wilson *et al.*, 2005). Fundamentally, the approaches used for sentiment analysis can be divided into two categories, the machine learning approach, and the lexicon-based approach.

1. Machine learning approach

   This approach is a supervised learning approach. It is referred to as a training process that teaches an agent to classify inputs to outputs. Once a qualified set of training data labeled with sentiment values is learned by the algorithm, the sentiment analysis on the corresponding domain data will provide promising results (Pang *et al.*, 02).

2. Lexicon-based approach

   This approach is typically an unsupervised approach. It is a rule based on features provided by predetermined a sentimental lexicon score to estimate the polarity whether it is positive or negative. These approaches can function without any reference corpuses and preceding trainings. Sentiment lexicons usually correctly estimate the generic polarity of terms in a way that they do not take domain information into account (Zhe, 2010).

Opinion words are words that are used to state positive or negative sentiment. Words that express desirable feeling, such as great or excellence have a positive polarity while words that express undesirable feeling, such as bad or awful, have a negative polarity. The dictionary of opinion words to identify a sentiment and its orientation is called “opinion lexicon”. Some of the interesting opinion lexicons is SentiWordNet and Pageranking WordNet.

SentiWordNet is one of public linguistic resources that are assigned to each synset, or synonym set, of WordNet with three numerical scores: positivity, negativity, objectivity (Esuli and Sebastiani, 2006). Pageranking WordNet is an applicability of a random-walk model to the determination of WordNet synsets in terms of how strongly they possess a given semantic property (Esuli and Sebastiani, 2007).

Sentiment analysis on the micro-blog has received a scholar intention recently. The micro-blog contains important information that states positive or negative feeling in very limited space, for example, “The htc hd2 is awesome, the 1GHz processor really helps.”

However, it is not always as easy as identifying “I love android” or “He hates iPhone.”. The language people speak today is complicated and full of slang, ambiguity, sarcasm, idiom, and irony as in the following example, “HTC battery die in 15 minutes but take a year to charge. I'm so lucky!”.

Early research on this issue may be the twitter sentiment classification using distant supervision (Go *et al.*, 2009). This work showed that machine learning algorithms have accuracy above 80% when trained with emoticon data and SVM outperforms other classification while unigram model surpass bigram and parts-of-speech (POS) features. Moreover, there is a study confirm that POS features are not useful for sentiment analysis in the micro-blog domain (Kouloumpis *et al.*, 2011).

Later interesting work is the combination of both the lexicon-based and the machine learning-based approaches. In this study, the researcher exercises a lexicon-based approach to perform entity-level sentiment analysis, which can give high precision but low recall. Then, to improve the recall and the F-score, machine learning-

based approach is also applied to identify an opinion automatically by exploiting the information in the result of the lexicon-based method (Zhang *et al*., 2011). To further study some commercial product available in the market, please check in Appendix C.

Business Intelligence

The Business Intelligence (BI) is a socio technical concept that transforms raw, meaningless data into useful and actionable information (Evelson, 2007). It mainly refers to processes used in identifying, extracting, and analyzing business data to support management activities especially in their decision making tasks. Business intelligence implementations encompass many different technologies, including OnLine Analytical Processing (OLAP), Extraction Transformation and Load (ETL), Data Quality (DQ), combined with organizational capabilities that allow corporate to use information to support business their processes.

A. Business Intelligence history

Previously, Information Systems (IS) were focused on solutions to help performing routine tasks as fast as possible to enhance corporate performance. Later, the capacities of processor and data storages became more accessible and flexible. The use of Information Systems expanded from purely an automation perspective to "informating", empowering ordinary working people with the overall knowledge to make them capable of critical and collaborative judgments (Zuboff, 1988). Organizations started migrating to complete Business Intelligence environments so that they could have a single version of the truth through the use of the cross-organizational data, provided by an integrated architecture (Eckerson, 2003; Negash, 2004). In recent years the Business Intelligence systems have consistently been rated as one of the highest priorities of IS and business leaders (Evelson 2007).

Business Intelligence 2.0 is technology in conjunction with social web trends that refers to a tool for enabling businesses to obtain dynamic query of real-time information. It emphasizes the importance of finding meaningful knowledge from an online

community or social networks to create meaningful data measurements. This real-time user data measurement can reveal consumer insight to check whether the business is growing or shrinking in any economic situations or needs to branch out in a new direction. For example, BI 2.0 can retrieve these kinds of information from Twitter: opinions about the products/services, the continuous flow of consumer commentaries, and event sequences.

As shown in Figure 2.7, (Constantinides and Fountain, 2008) presented that Kotler's Stimuli and the Response model have been changed because the usage of Web 2.0 or the Social media. Different from conventional market, consumers are not only influenced by the traditional marketing (A), but also the irrepressible personal influencer, such as culture, economic, and legal (B). Currently, consumer includes two more factors to the model. The first is the traditional web experience (C), which fundamentally represents the controllable online activities, provided by the producer. The second can be called "social media experiences" (D), which is plentiful beyond the control of marketer.

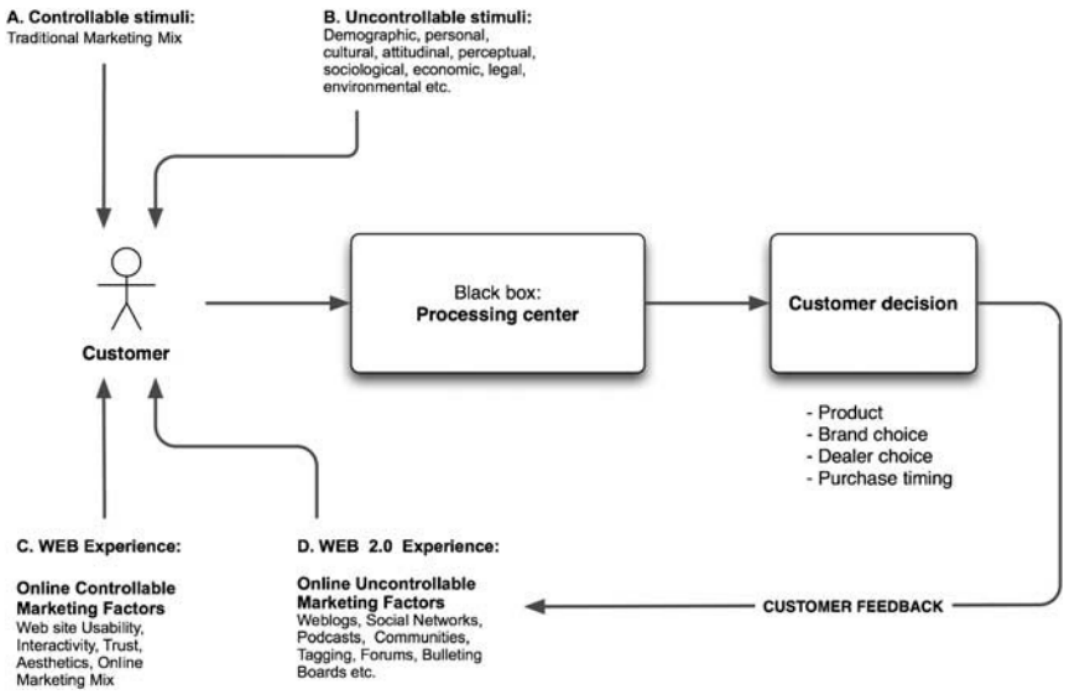

Figure 2.7 Kotler's Stimuli and Response model

B. Business Intelligence for Consumer Insight

The consumer insight is the study about classifying the consumers and their thought or feelings (Stone *et al*., 2008). It may be the revolution from a direct marketing to a database marketing to the customer relationship management or the customer experience management. Enterprise can utilize this understanding to support from changing their marketing strategies to improving their operations and interactions. The suitable consumer insight study is the foundation of decent Customer Relationship Management (CRM). Therefore, enterprises will be able to position themselves as they want it to be to meet customers' or stakeholders' needs.

Insight is not conscious behaviors or thoughts of consumers but most are affected by various external factors from the state of economy and society to the way a brand is marketed. Marketers can collect this information from both criticisms and compliments for further study. Some tasks can achieve from a market research while some need mining from a customer database, feedbacks from sales and customer service staffs.

Although measuring the customer satisfaction has become very significant, setting staff targets using information based on the conversation among consumers can stifle enterprise's creativity and also limit the innovation. In a large corporate, the measurement will mislead by reducing cost and increasing control. This focus has become internal; thus, it leads staffs and the entire organization away from first considering the customer. As a consequence, they fall in the customer loyalty and the overall satisfaction. The consumer insight process helps changing the understanding of marketers about consumers, staff and their organizations, rather than supports a counterproductive way of thinking about them.

Applications for mining large volumes of the textual data for marketing intelligence can be categorized into three types (Glance *et al*., 2005) as follows.

1. Early alerting - informing subscribers when a rare but critical, or even fatal, condition occurs.
2. Buzz tracking - following trends in topics of discussion and understanding

what new topics are forming.

3. Sentiment mining - extracting aggregate measures of positive vs. negative opinion.

An Opinion Observer is an early framework for analyzing and comparing consumer opinions of competing products in the market (Liu *et al.*, 2005). The system provides a single glance of its visualization. Users are able to clearly see the strengths and weaknesses of each product in consumers' mind in terms of various product features. For a product manufacturer, the comparison enables it to easily gather the marketing intelligence and product benchmarking information. Second, a new technique based on the language pattern mining is proposed to extract product features from positive and negative in a certain type of sources.

CHAPTER III
METHODOLOGY

This chapter consists of explanation on the research methodology to be followed for this research work. The researcher adopted both qualitative and quantitative research approach to conduct this research. The work was carried out in four stages. First, a focus group discussion was conducted to identify the need of a new marketing tool on the micro-blog. Then, in the second stage, in-depth interviews were held to find the direction of the software prototype development. Third, a software prototype was implemented for the experimental study. This stage is divided into several procedures that will be outlined in more details under this chapter. Last, a qualitative study was carried out again to closely examine more in details of the software prototype by a focus group discussion with another group of experts.

The initial focus group discussion

In the beginning of this research, the need of an innovative marketing tool on the micro-blog was not quite obvious since this technology was quite new and there were not many people using it. To build a confident in serving the need of the current market, the researcher obtained an idea to gather information from the top level executives from various industries first. A focus group discussion was seen as an effective way to generate the descriptive information that was required as a preliminary step, providing the background information and defining the direction for the field-study.

A focus group discussion is used to acquire attitudes, perspectives and knowledge of people as in face-to-face interview but the significant advantage is the collaborative wisdom from many points of views of several participants in a way that they would not be easy obtainable in responses to other methods. Therefore, the first focus group discussion with executives was conducted to collect information of several perceptions related to the micro-blog from the top level management. The feedback

provided by the focus group participants were expected to clarify differentiated aspects of this research domain and to identify key elements in the market requirement on OM on the micro-blog.

The recruitment criteria were the executives who have experiences in their industry more than 10 years and hold a senior position in their organizations. Participants were recruited through a variety of means, such as through references, and via snow ball technique.

A topic guide was developed that sought to elicit the discussion on following objectives.

- Objective 1: To understand the use of the micro-blog in the organization
- Objective 2: To determine the need of the micro-blog for the market research
- Objective 3: To gather requirement for the software prototype development
- Objective 4: To find business opportunity for the market research service on the micro-blog

The session was facilitated by researchers and advisors. The discussion guide, as can be found in Appendix A, was prepared to arrange an outline for the moderator to ask and inquiry interested issues. The moderator conducted the focus group using this discussion guide included explaining objectives and method of the discussion session, encouraging each participant to express their considerations even if it differed from others, and ensuring that participants were aware that there are no right or wrong answers in the session. As shown in Figure 3.1, the discussion was conducted in a comfortable room. Broad, open-ended questions were posed. An expression of interest was distributed across the participants to attend a focus group to discuss about the micro-blog roles for purposes of a marketing tool. The written transcripts, supplemented by the video recording, formed the basis for the model of the innovative marketing research in this study. The key themes finding are illustrated with anonymous quotes from different participants.

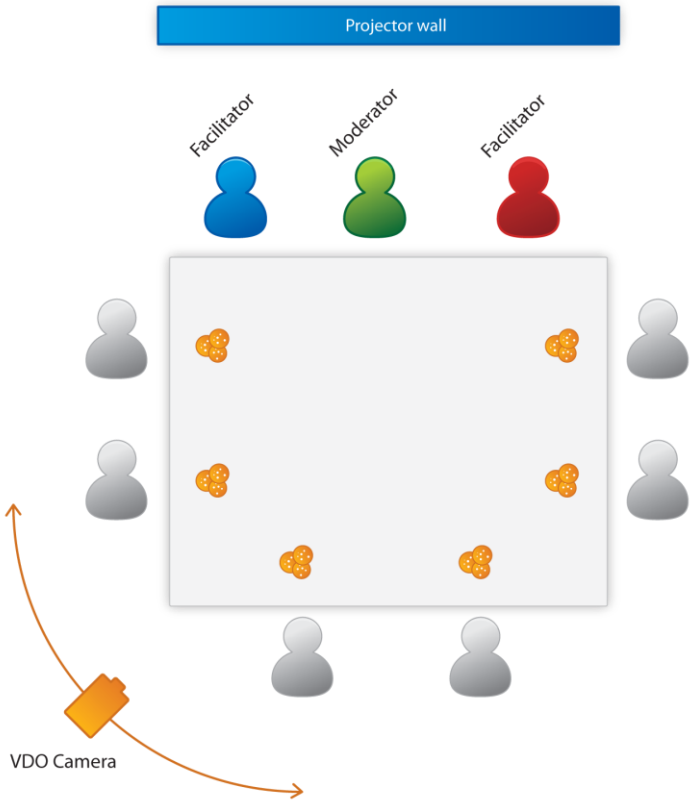

Figure 3.1 The focus group discussion layout

In-depth interview

The purpose of conducting in-depth or face-to-face interviews was to find out more specific requirement for the software prototype implementation. The interview is an adaptable and flexible way of finding the qualitative information. This approach provides a unique research opportunity to allow interviewees to clearly deliberate their opinions, choices and motivations (Lindlof and Taylor, 2002).

The researcher used the semi-structured interview that combines specific questions to gain information on specific topics plus open-ended questions to gain wider views on the current micro-blog usage and their prediction of the applications for the future. In this research, five in-depth interviews were conducted. The selection of the interviewees is an important process to obtain useful information and particular results. Five experienced experts who directly involved with the social media and/or the

consumer insight from various organizations were carefully chosen in order to gain the industrial as well as the academic perception.

At first, the researcher introduced the research domain, the problem definition and the research questions. Then, the interviewees shared their knowledge regarding the research domain. They also explained methods that they work on the marketing research both the traditional approach and the new approach. The interviews were recorded in the form of audio recordings with the consent of interviewees for further review and analysis. Moreover, the researcher also took notes during interviews.

After each session, the research carefully listened to the dialogue to expose information related to the research questions. For analysis purpose, the researcher prepared summaries of interviews for better understanding and removing unnecessary contents. The summarized interviews contain identified key points regarding the research questions.

Software experiment

In this section, the approaches and procedures used for solving the opinion mining task is described. The researcher developed software prototype to respond to the result from the initial focus group discussion and five in-depth interviews that had been conducted. (Hu and Liu, 2004) stated that one form of the sentiment analysis in product reviews is to produce a feature-based summarization. For this study, features of a product are attributes, components, and other aspects of the product, such as size, power, and display.

The research methodology differs from other researches in NLP field because the researcher specifically analyzed micro-blog messages using both the machine learning-based approach and the lexicon-based approach simultaneously. Moreover, this study applied visualization techniques in alignment with the OM concept to classify messages dealing with key success factors and failure factors of the product to help business decision support and planning.

In this study, the researcher proposed a solution, called the Micro-blog Sentiment Analysis System (MSAS), to analyze the opinion on Twitter. As (Jansen, 2009) mentioned that all micro-blog share a set of similar characteristics

1. short text messages
2. instantaneous message delivery
3. subscriptions to receive update.

As a result, the researcher assumed that the results should be valid to other micro-blog platforms aside from Twitter. The proposed system, the MSAS, could determine positive or negative sentiments on product features; then, aggregate the results to produce a summary for the users.

As illustrated in Figure 3.2, the MSAS processes can be separated into two phases: the Preparation phase and the Analysis phase.

A. Preparation phase

The Preparation phase is the stage that all required data, including a posts, model, a relevant lexicon, are arranged on a specific domain as follows.

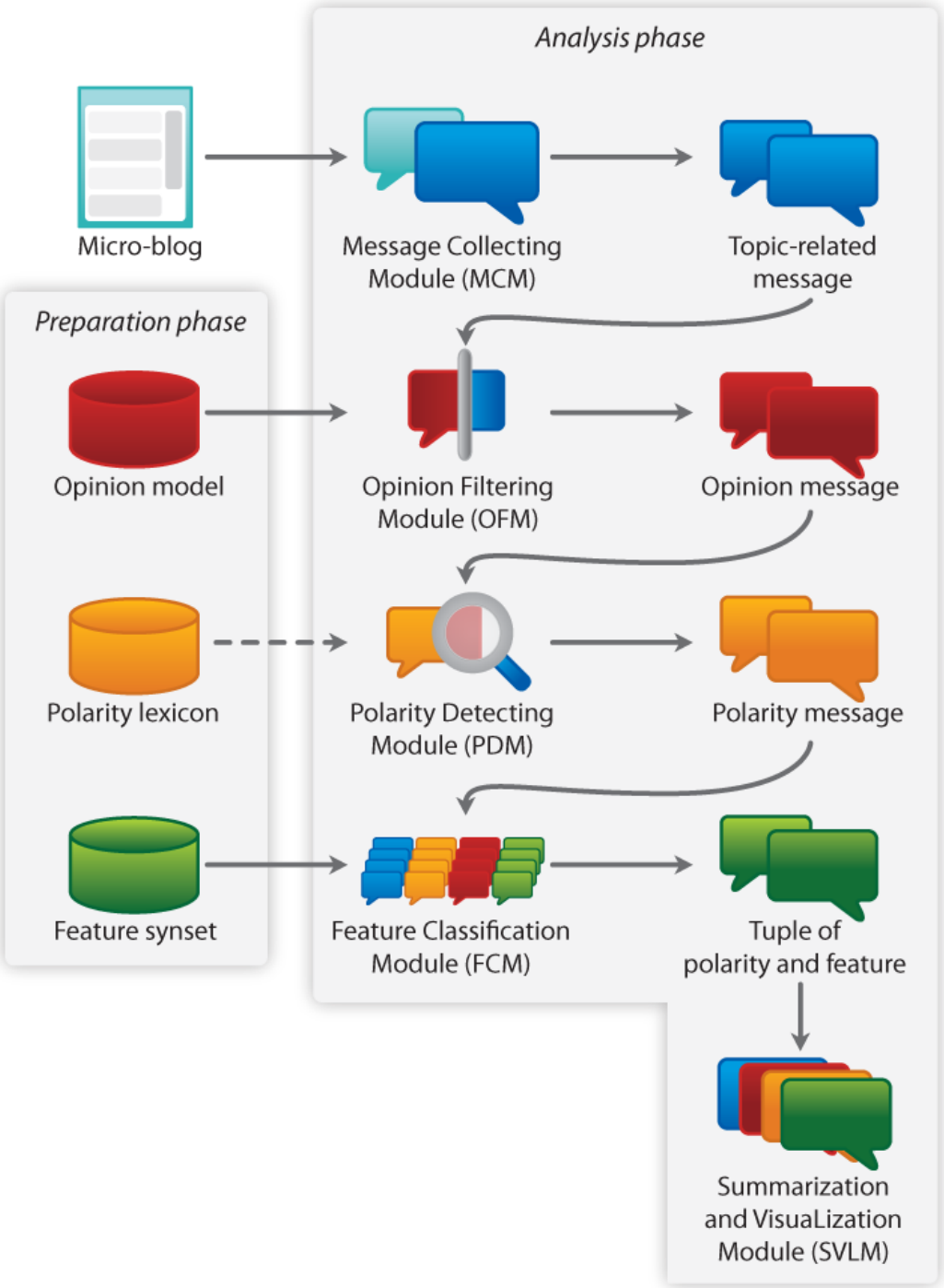

Figure 3.2 The Micro-blog Sentiment Analysis System (MSAS)

1. Micro-blog

This is the input source to the system. As mentioned earlier, Twitter can be a good representative of the micro-blog since it provides

tremendous amount of data that is available as a public timeline for easy gathering. As illustrated in Table 3.1, the researcher used the domain of smartphones as a case study. For example, record ID 1 is created on Auguest 28th, 2010 time 00:49:42. Its ID Twitter is 22285927850. The application is twidroid.com. The message is "I love this Samsung Galaxy but if u get one dont plug it in straight away to fully charge the battery. Instead let the battery run out..." by user Rafayells.

Table 3.1 Examples of collected tweets stored in database

id	created_at	message_ status_id	source	text	user
1	2010-08-28 00:49:42	22285927850	twidroid	I love this Samsung Galaxy but if u get one dont plug it in straight away to fully charge the battery. Instead let the battery run out...	Rafayells
2	2010-08-27 22:05:01	22274155397	web	Motorola Milestone has the 'best mobile display in the world'. Ref: displaymate.com	teenbreakthru
3	2010-08-28 02:32:26	22292491844	<a href="http://tinyurl.com/tweetcaster" rel="nofollow">TweetCaster</a&	None of the three cameras I can videotape with focus on close up items. But my iPhone 4 does. How lame is that?	jenfrtheblock

2. Opinion model

This model is created based on the classification algorithm. The model is used for filtering the opinionated posts from the non-opinionated

ones. Table 3.2 shows some examples of tweets classified as opinion and non-opinion posts. Moreover, the opinion message can be classified as positive or negative. The sentiment of each tweet will be further analyzed using polarity lexicon.

Table 3.2 Examples of opinion and non-opinion tweets

Type	Sentiment	Message
opinion	Positive	wow that's awesome! Enjoy customizing. :) Nexus one is probably the father of customizations with several types of roms available
	Negative	My Nexus One is not responding well to location change for mobile network signal change. Need an app to manually refresh network.
non-opinion	N/A	Look for Verizon's new release of the Motorola Devour as well as an HTC Hero coming soon. AT&T to finally introduce Android OS!!!

3. Polarity lexicon

It is a dictionary used for the lexicon-based sentiment analysis approach. The researcher obtained the opinion lexicon from SentiWordNet 3.0 (Baccianella *et al.*, 2010). However, this study discarded objective terms since it does not contain positive or negative polarity. Table 3.3 shows some examples of terms from SentiWordNet 3.0. In this table, the POS refers to Part of Speech; the pair (POS, ID) uniquely identifies a WordNet (3.0) synset. The following letters are the POS notations:

n - NOUN

v - VERB

a - ADJECTIVE

s - ADJECTIVE SATELLITE

r - ADVERB

There are three scores involved with the calculation, which are positive score, negative score, and objective score.

objective score = 1 - (positive score + negative score)

The values PosScore and NegScore are the positivity and negativity scores respectively.

SynsetTerms are sets of synonymous terms with sense number written after the # symbol. The number notation comes from WordNet, which represents the rank in which the given word is commonly used. The rank#1 refers to the meaning of the rank of most commonly used. Lastly, Gloss is the meaning of that word.

Table 3.3 Data format of SentiWordNet v3.0

POS	ID	PosScore	NegScore	SynsetTerms	Gloss
a	01586866	0.75	0	pleasant#2	(of persons) having pleasing manners or behavior; "I didn't enjoy it and probably wasn't a pleasant person to be around"
a	01587077	0	0.875	nasty#1 awful#3	offensive or even (of persons) malicious; "in a nasty mood"; "a nasty accident"; "a nasty shock"; "a nasty smell"; "a nasty trick to pull"; "Will he say nasty things at my funeral?"- Ezra Pound

4. Feature synset

This is a collection of multiple product or service features that companies would like to study. Since different customers may refer to each feature differently, the researcher had to group them into feature categories by studying the product specification across all manufacturers' websites. Table 3.4 shows some examples of feature categories and feature terms for smartphones. For example, if the post contains the word "case", it should be classified in the Accessories category.

To prepare features for other studies, the researcher has to define category first, and then list and group each feature to the

category.

Table 3.4 Examples of smartphone feature categories

Category	Features
Accessories	case, headset, cable, dock, remote, mic
Application	app, application, game, twitter, facebook, browser
Camera	video, photo, camera, vga, led, focus, megapixel, mp
Capacity	capacity, disk, storage, ram, rom, memory
Network	2g, 3g, connection, wifi, wireless, cellular
Power	power, battery, hour, recharge, standby, charge
Screen	display, ppi, pixel, capacitive, resolution, monitor, lcd
Sensor	accelerometer, sensor, proximity, vicinity
Size	height, width, mm, cm, depth, weight, ounce, gram
System	system, cpu, speed, processor, GHz

B. Analysis phase

Once all data and related lexicons had been already prepared, the Analysis phase is the examination stage composed of four related modules as follows.

1. Message Collecting Module (MCM)

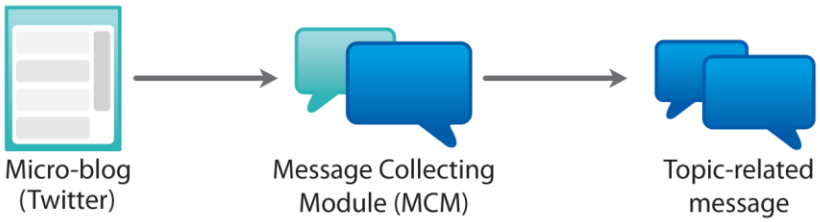

Figure 3.3 Message Collecting Module

As shown in Figure 3.3, this module utilizes Twitter Search API to collect tweets for the study. Twitter has an Application Programming Interface (API) for programmatically accessing tweets by the query term (Twitter website, 2011). The search result is defined as anonymous

without authentication. At the time the researcher collected data, Twitter API has a limit of 100 tweets in a response for any requests. Data can be retrieved as XML or JSON formats. For smartphones, the researcher uses query terms by the following related keywords, such as iphone, nexus, htc, motorola, nokia, blackberry, and palm.

The Twitter Search API is a channel to obtain tweets that were interested in this study. The results are not the complete index of all tweets, but instead an index of recent tweets that match a specified query between six to nine days. However, Twitter will not allow the Search API to find tweets older than about a week. Therefore, the search results were collected and accumulated in the local database for further study.

As of April 2010, the Twitter Search API provides an option to retrieve the "popular tweets" on the top of real-time search results. Then, as of November 2011 the Search API returns Twitter user IDs that match the Twitter API. Developers should watch carefully since Twitter API has been continuously modified.

2. Opinion Filtering Module (OFM)

As shown in Figure 3.4, this module classifies a given tweet as "subjective (opinion)" or "objective (non-opinion)" using a machine learning approach. First, the study started by collecting training data and labeling tweets manually as opinion or non-opinion. Then, the researcher performed the following tasks: feature selection and classification model construction. According to the primary study of this research, the researcher found that the best performance for filtering opinion tweets is the use of the Support Vector Machines (SVM) with the Information Gain (IG) feature selection. The model yielded the accuracy equal to 84.5% (Wilas *et al.*, 2011).

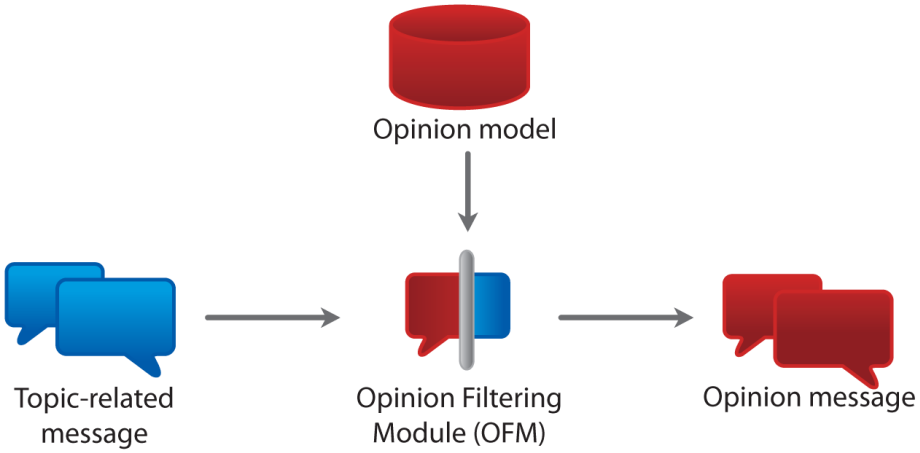

Figure 3.4 Opinion Filtering Module

As illustrated in Figure 3.5, the proposed method is called the "Opinion Filtering Process" which consists of three main steps as follows.

a) Text processing to clean input micro-blog posts. This step will clean the data source with four standard processes since Twitter contains very casual and informal languages.

- tokenization – to segment each post by observing the word delimiters, such as space, and punctuation marks.
- stopword removal – to remove stopwords from the bag of words by looking up a stopword dictionary. The stopword removal helps reducing term dimensions by cleaning up the unmeaning terms.
- link removal – to remove address and external link, since they are not useful in detecting polarity and features.
- stemming or term normalization – to change the form of verb and adjective back to the base form, and replace the abbreviation with its meaning. For example, the verb "talking" is replaced with "talk".
- slang handling – to replace a sequence of repeated characters by detecting a sequence of three or more

characters, for example, the word “goooooooood” is reduced to “good”

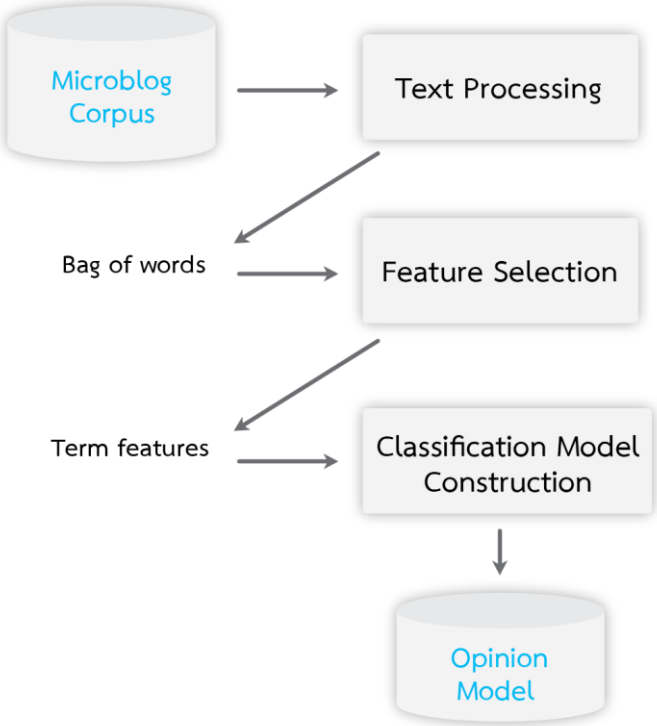

Figure 3.5 Opinion Filtering Process

b) Feature section for representing and selecting terms in each post. Example posts are shown as follows.

- Opinion post

  “Wow… The new iphone has video calls... That's some #Jetsons ish lol ~TweeDyBoyFitted~”

- Other post

  “More Details on Those New HTC Codenames http://bit.ly/cNEbLM (via @FlyScreen)”

c) Classification model construction as a knowledge management tool to filter the opinion post from others

3. Polarity Detecting Module (PDM)

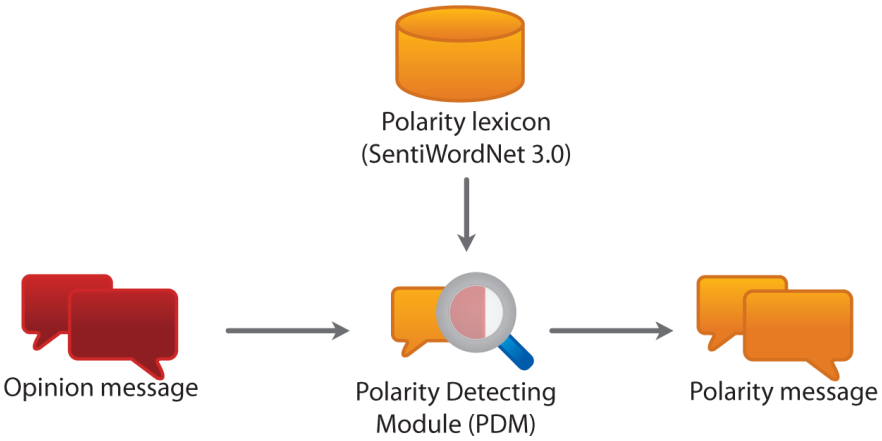

Figure 3.6 Polarity Detecting Module

As shown in Figure 3.6, this module determines the polarity of a given message as positive or negative. It could also indicate the strength of the sentiment words. The module calculates polarity score for each tweet by averaging the scores of each sentiment words found in the tweet following these rules.

a) initialize the total polarity score: s ⟵ 0

b) check each token with SentiWordNet

   if token is positive, then s ⟵ s + w; where w is score of sentiment word

   if token is negative, then s ⟵ s – w

c) if tweet contains negation, then s ⟵ -s; where negation are the word “no” or “not” that will twist the overall tweet polarity

4. Feature Classification Module (FCM)

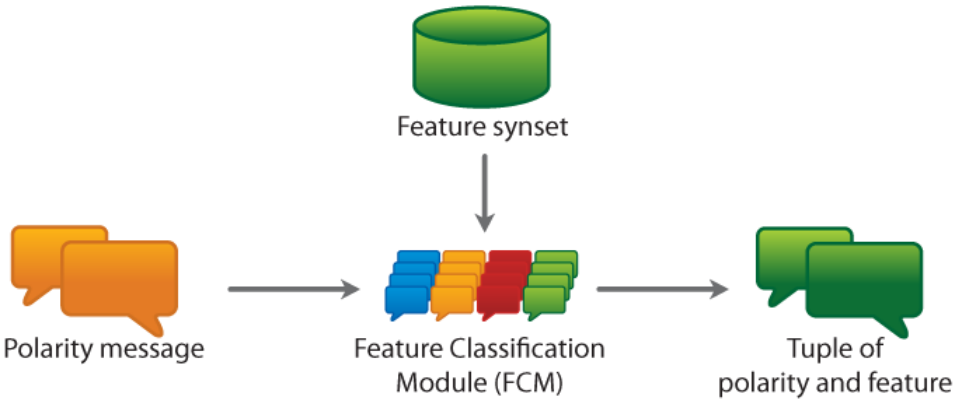

Figure 3.7 Feature Classification Module

This module is responsible for outlining the brand or the product feature of each tweet. For example, “Twitterific is pretty great. I do all my tweeting through my iPhone”. This tweet is classified as follows.

a) twitterific, tweeting ⟶ feature

b) pretty ⟶ sentiment (score = 0.2083333333333333)

c) great ⟶ sentiment (score = 0.1069303914658766)

d) iPhone ⟶ brand

5. Summarization and VisuaLization Module (SVLM)

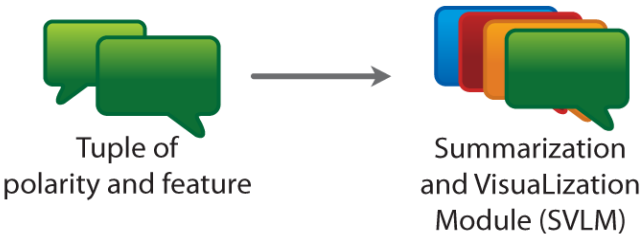

Figure 3.8 Summarization and VisuaLization Module

In the last process, as illustrated in Figure 3.8, the MSAS will provide the overview of the sentiment polarity for each product features that people post though the visualization of feature-based summaries of opinions. Three different visualization techniques were adopted as follows.

a) Radar Chart – to compare many properties of one interesting issue in one graph. Obviously, the greater the area covered by the graph, the greater the overall value.
b) Bar Chart – to highlight separate quantities, especially the differences between quantities in several categories of data.
c) Line Chart – to show the pattern of changes of a value. Moreover, the chart can show multiple pieces of data by using multiple lines in the same graph.

The final focus group discussion

Lastly, the researcher conducted a focus group discussion again to verify the result of the software prototype. This time, the researcher would like to find out what people in the industries think about the benefit of marketing research on the micro-blog and to assess the impact of this software prototype.

Prior to conducting this focus group discussion, the researcher had developed a self-administered questionnaire with a focus group discussion guide, as can be seen in Appendix B. Like the first focus group discussion session, this discussion guide should be able to promote straight conversation, and provide promptly replies to motivate participants to exchange their ideas and reengage whenever the discussion paused. The self-administered questionnaires comprised both qualitative and quantitative questions were distributed to the participants before the session.

An important goal in selecting participants for this focus group was to include members of people who must have expertise in both academic and the professional fields. Nine Participants were recruited through a variety of means, such as through references, and via snow ball technique.

A topic guide was developed that sought to elicit the discussion on following objectives.

- Objective 1: To evaluate the concept of the Micro-blog Sentiment

Analysis System

- Objective 2: To ask for the acceptance of the software prototype as an innovative marketing tool
- Objective 3: To look for any other suggestions that experts may have

For evaluating our software prototype in Objective 2, the Technology Acceptance Model (TAM) in employed to the study. As illustrated in Figure 3.9, to determine an individual intention to use a system, the researcher have to figure out Perceived Usefulness and Perceived Ease of Use that lead to the attitude toward using the system (Davis, 1986).

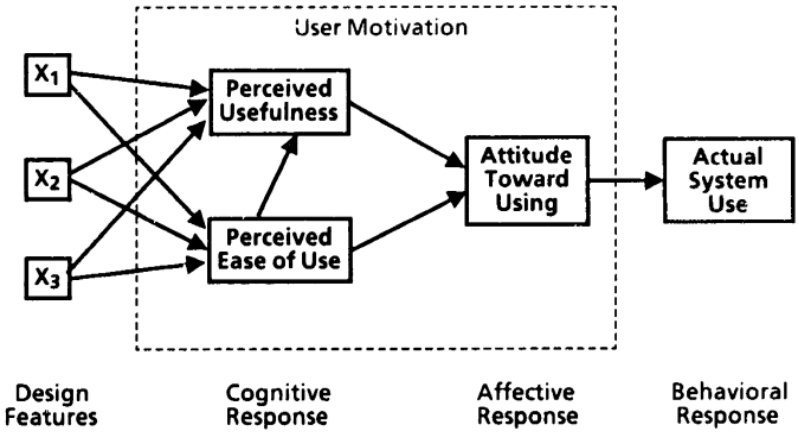

Figure 3.9 Technology Acceptance Model

Again, this focus group discussion was conducted in a relaxed manner. Participants were assured their anonymity would be maintained. Again, the session was facilitated by researchers and advisors. The procedures are as same as the first focus group discussion that was conducted.

CHAPTER IV
RESULTS

This chapter, which describes the results of the study, is presented in four sections followed the methodology in Chapter III.

1) The initial focus group discussion

2) The in-depth interview

3) Software experiment

4) The final focus group discussion

The initial focus group discussion

A. Venue and schedule

The discussion was conducted on 23rd June 2010, 13:00-16:00 at 14th floor Chamchuree Square, Technopreneurship and Innovation Management Program, Chulalongkorn University.

B. Participants

Participants composed of seven experts in various fields as follows.

1 Software development

1 Mobile phone manufacturer and distributor

1 e-commerce solution provider

1 advertising agency

1 department store

2 property development

C. Result

There are four objectives to be accomplished as follows:

Objective 1: To understand the use of the micro-blog in an organization

- All participants understand and trend to exercise the micro-blog for a personal use. Tough none participant had utilize the micro-blog for business use seriously; two of them had planned to do so.
- One participant expressed that the use of the micro-blog for consumers is very useful for listening to opinions of other consumers. Currently, consumers do not believe opinion from producers, but they trust the fact from producers.
- For the business use of the micro-blog, it may divide into two types, which are the internal use and the external use.
- All participants agreed that the micro-blog for the external use is for communication and public relationships to the target audiences. This tool is very effective and not complicated. However, there are some concerns as follows.
    - Negative feedback handling
    - Corporate message controlling
    - Number of the micro-blog users is quite low and not yet significant
    - Analysis result measuring
- One participant explained that the micro-blog can be employed for the internal use for collaboration among teams. Anyway, any negative feedbacks are very useful for corporate to understand other people attitudes. The earlier the company knows the problem, the faster they can solve it.
- One participant also declared the ban of the micro-blog use in a corporate policy due to over time spending in office hours.

Objective 2: To determine the needs of the micro-blog for market researches

- All participants agreed that the micro-blog is a good channel to express idea. It is very useful for corporation to check consumer attitudes according to these factors:
    - People dare to express themselves

- Information technology is more advanced: affordable and fast
- Many devices in the market, many choices for all preferences

- Four participants concerned about the volume of data collected from the micro-blog whether it is too small or not. Although there is potential to increase constantly, the target customer may not adopt the micro-blog in their life yet.
- Three participants thought that the identification of people who declare opinions is important. So, the problem is about checking whether opinions really come from customers.
- Two participants aware of the current written language which is quite difficult to understand for a machine; even for a man.
- One participant suggested that the micro-blog for the market research should be used supplemented with the traditional approach. It's an alternative approach, not a substitute one.
- One participant thought that the result from this research on the micro-blog could lead the target group to participate in the later in-depth interview.
- One participant considered about methods to explain the market research on the micro-blog for the ordinary people.
- Two participants realized about influencing effects when someone speaks and other believe without deliberation

Objective 3: To gather requirement for a software prototype development

- All participants agreed that the market research on the micro-blog is able to investigate everything depends on the proper application for example:
  - R&D planning to maximize the production
  - CRM to listen to the customer voice
  - Competitor monitoring
  - Marketing for finding a segment of target customers
- One participant suggested that the application must be user friendly.

- One participant concerned about the catchment and sampling. The approach that followed statistic methods will yield high validity.
- One participant questioned about beneficial for the company when they apply the result whether the collected information can represent target audiences, or not.
- One participant concerned about the privacy issue.
- Three participants realized about the limitation of Thailand infrastructure.
- One participant proposed to consider other sources for study, such as e-mail, or SMS.

Objective 4: To find business opportunity for the market research service on the micro-blog

- All participants were interested to use this kind of service and did not want to develop in-house because they realize the difficulty of this software implementation. Price is upon for what they are using.
- One participant suggested that the company that offers the tradition market research to their customers should offer with this kind of service.
- Two participants believed that this kind of service requires high confidence of the service provider. It should be setting up by a trustful organization and open aggregate results to the public for the benefit of the entire industry.
- One participant suggested to concern about the first group of customers which may drive this service to be able to commercialize as the real business.

In-depth interview

A. Schedule and participants

In second step, interviews were conducted during 30th August to 2nd December

2010. Initially, interviewees were briefed with a short explanation about the purpose of the research on the phone or e-mail. Then, the research went to interview at the place and the time that interviewees were convenience. The interview sample composed of five experts in the fields as follows.

1 Software development

1 Writer

3 Media agency

B. Result

In summary, each interviewee provided their opinions as follows.

First interviewee described that one way for using consumer insight is for communication in advertisement. It is the concept verification in deep down into the real needs of consumers.

- The consumer insight research is mostly in the tail of a supply chain since the product is already produced and producers require searching for the channel that they can put some product's attributes to the target group or beat the competitors.
- For the product pretest, this test usually relies on some images or some words whether it is suitable or not.
- Mostly the interviewees conduct a focus group discussion for 5 groups; each group contains 8 - 10 people, to cover the assumption. It would take around 1-2 hour per group. There are many methods, such as story board telling, responding the specific question, and scoring on the questionnaire while discussing.
- The result of this research is confident in some ideas or words. Respondent profile is required to make sure that they are the target group whom the interviewee would like to study. The researcher must do whatever methods to persuade respondents to talk about their opinions. Researcher must find out

the strongest point in the consumer thought for communicating back to them. However, sometimes the focus group discussion has a problem about guiding too much.

- Final result that the interviewee provides to customers is a report regarding methodology and summary headline for advertisement.

Second interviewee explained that the social media can be used for both offensive and defensive. Offensive is for representing themselves as a brand straight forward. Defensive is for acquired information about the brand who is talking about us.

- Though the online research is easy, fast, and safe cost, little research uses this method because of its limitation as follows.

  - Digital divided. Only a group of people uses the Internet. Online research is suitable for people who live in the urban area.

  - Confidential of the product. If it is the product-concept testing, it is hard to control the confidential of information.

  - Face-to-face is more reliable since researcher can communicate to target audience directly since they know to whom they are talking.

- Many times open-end questions cannot gain deep information since the respondent cannot really understand the point. Therefore, the interviewer must provide some hints or choices to clarify what is needed.
- Three factors in getting consumer insight

  1) good design research
  2) proper stimulus
  3) accurate analysis

- For corporate, sometime they use social network too much to advertise over the truth. Finally, when the truth appeared, they have to accept and find the way to correct the problem. Brand must be honest to consumers.
- Interesting output for the online research should be tracking research with

the regulated time, and also be able to predict the trend in the future.

Third interviewee said that though customers do not buy today, they may talk about what they want. If marketer can collect and process these data, it will be very useful for R&D to answer the customer needs. Besides, the marketer will be able to check market competition, especially in the high competition market, such as bank and telecommunication.

- There is no general influencer. Influencers are depended on types of industries. The marketer has to cluster data and weight it first, then prioritize the data on the degree of influencing.
- The marketer would like to have mechanism for data capturing and bring it back to the call center. If there are any complaints that the business cannot handle, it will bring bad experience to the customers. Thus, the crisis management is important since what happen on the Internet is going on viral so fast.
- Time to process is another important factor. If it takes too long time, the summary that the marketer received will not represent anything which happens in the present. Information on the Social network is the snapshot of time. The marketers can use for their improvement area, for example, Hong Kong Government uses to analysis citizen sentiment for issuing the management policy in different period.
- Corporate should integrate the result of the sentiment analysis to legacy system, such as CRM.
- Many departments can utilize this information, such as marketing, customer support, PR, R&D, business development, and brand management.
- While above the line activity is reduced constantly, below the line activity is an alternative way that marketer would like to invest in. Therefore, information about consumer insight will help marketer a lot.
- Thailand lacks of market readiness and need more education. Everything is

hype but most corporates are not serious with this kind of tools.

- Problem for the sentiment analysis tool are as follows.
    1) It's not provided completed loop for analysis. The marketer needs data that has close relationship to people but there is limitation of data accessible in the social media. Most social network has privacy rules to limit the data accessing which is not a technical problem, but it is related to the service policy.
    2) Non 100% automated system. However, there should be components between the automatic process and a judgment by experts. Analysis process still requires experts to transform the qualitative data to the quantitative data.

Forth interviewee mentioned that voices from the micro-blog are real sentiment from consumers, which may cause good or bad effect and may cause a serious problem once the speaker is the influencer towards people, especially the issue that may be easy to viral.

- The industries that are able to study from the micro-blog are Food, Travel, Lifestyle and Technology product.
- People love to talk about this issue in the micro-blog. If Brands would like to listen for events aroused in the micro-blog, they have to be in the mentioned industries above.
- People who are not online will have idea and lifestyle in using products or services different from those who are online.
- The micro-blog can be media for people to build their personal branding. Everybody can show their identity by posting in what they are interested and stimulating others follow. They might be some famous people or just an ordinary people who talk about their experiences and can persuade people to become a huge network. Therefore, the intensity of influencing may measure from a number of followers.

- The reliability of the micro-blog may be 70:30. Since the influencers do not want to reduce the number of their followers, what causes other bored or causes timeline dirty with too much post will be avoid. However, when brands know more about the micro-blog and study more to know who influencers are in specific issues, messages in micro-blog will be manipulated in the way that brands want. Although some influencers are paid to post some advertising, presently consumers are clever enough to know what is advertising or not and the confidential of that influencer will be reduced. However, the issue that be talked from many people simultaneously will be more reliable than only one influencer talk.
- The first mover in each industry who involves in the social media will be more advantages than others who came later. There are many brands which are not in the top list of the market, but they can build their brands over the Internet.
- The micro-blog is not enough to represent a voice of a consumer. In Thailand, there is more information from webboard, blog, and Facebook than the micro-blog. Besides, most users use Twitter to build identity. They are not share what they feel in Twitter.
- The sentiment analysis tool should have these characters:
    1) Top ten keyword for like and dislike of product
    2) Graph presentation
- The business model for this software may be separated into 3 tiers:
    1) Sell to a big corporate and customize software as they want. The revenue will be both licensing and outsource fee.
    2) Sell to an agency by separating layer of results
    3) Sell to the public by subscription period

Fifth interviewee explained that the micro-blog using for corporate may be separated into 3 types:

1) Passive - waits for someone mentions about the brand
2) Active - joins the conversation in the community to ask for a feedback from each group
3) Proactive - listen more than mention to the business directly, but aware of proximity keywords related to the business.

- In Thailand, the interviewee believes that business listens to voices of consumers via the micro-blog, but they have not used the sophisticated tool. Mostly business still uses free services, such as search.twitter.com, or hootsuite.com
- The result from these tools provides raw data that who talk about keyword that corporate interested in, but there is no analysis to the level of attitude of those messages. Therefore, the marketer cannot study the behavior of consumers from this approach.
- Factors that should be considered in measuring the consumer interest are listed below.
    1) Number of speakers or messages
    2) Group of speakers
    3) Behavior of those speakers
    4) Time that they talk
- Purpose of the analyzing tool
    1) To understand the speakers who is talking about the focusing brand
    2) To further study of behavior that is positive or negative factors for consumers
    3) To be able to recommend or offer related products to consumers
- Reliability of data
    1) Tool should check the consistency of speaking
    2) Tool should separate between real consumers or spam
    3) Tool should find out who is an influencer using constraints as follows.
        - ratio between the follower and followings is low

- many people mentioned to
- ratio between tweet and retweet is high

- Privacy violence will be an impact number of the micro-blog users. People who never use will not turn to be a user. People who do know about the privacy setting will do it. People who do not care will not do anything. However, in the near future, people will be more careful about what they talk on the Internet because they will know that it will affect their lives or works.
- Business model that may suit sentiment analysis tool
  - Freemium: provides basic service for free and charges for some additional features
  - Free trial: to test the result of the tool first

Software experiment

A. Data set

In this study, the experiment is performed on the domain of smartphones since there are enormous number of conversations and thoughts related to smartphones on Twitter. Smartphone is a mobile telephone with extensive properties, such as a personal digital assistant, an Internet browser and other applications. Once the technology progress has increased the smartphone capabilities at an affordable price, the global smartphone market is incessantly growing. Smartphone becomes a part of many people daily life. As a result, tweets related to how people talk about the smartphone are captured in these experiments. All tweets from the public timeline spanning during March 2010 to June 2010 are collected via Twitter Search API and initially filtered using the following criteria.

1. Tweets must contain related keywords associated with smartphone brands as shown in Table 4.1
2. Duplicated contents are ignored.
3. Non-English tweets are also filtered out.

4. Tweets that contain more than one product will not be considered because it may be a comparative sentence which is beyond the scope.

Table 4.1 Related terms and keywords for smartphones

Operating System	Brand	Keywords
Apple iOS	iPhone	iphone
Google Android	Motorola, HTC, Sumsung, Google Nexus	motorola milestone, motorola droid, motorola cliq, motorola devour, motorola backflip, htc, samsung galaxy, samsung omnia, samsung wave, nexus one
RIM Black Berry OS	BlackBerry	bb bold, bb torch, bb curve, blackberry
Symbian OS	Nokia	nokia n, nokia e, nokia c
Palm OS	Palm	palm pre, palm pixi

B. Processing

As illustrated in Figure 4.1, the researcher collected around 1,000,000 tweets using the Message Collecting Module (MCM). However, the researcher randomly selected 100,000 tweets for the experiment to reduce the processing time and to ensure the evenly distribution of each brand along the timeline. The tweet collection is separated into 10 groups equally for each collecting time period. Therefore, 10,000 tweets are randomly selected from each group to put into the Opinion Filtering Module (OFM). To construct the classification model, the researcher manually annotated 600 randomly selected tweets, half of which contain opinion and non-opinion tweets. Based on previous study, the researcher found that for the opinion filtering model the Information Gain (IG) feature selection and the classification algorithm Support Vector Machines (SVM) yielded the best performance (Wilas *et al*., 2011). After processing 100,000 tweets with OFM, the system received the results of about 20,000 opinion tweets. Then, these tweets are sent to the Polarity Detecting Module (PDM) to analyze and assign the polarity scores for each tweet. The Feature Classification Module (FCM) will further extract product features and brand mentioned in each tweet. Finally, the

results will be displayed by the Summarization and VisuaLization Module (SVLM).

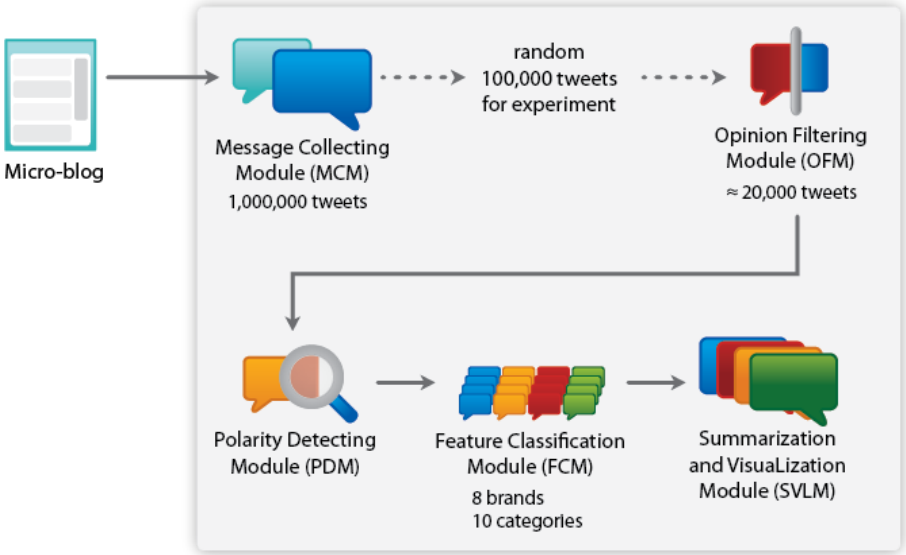

Figure 4.1 Overview of the experiment

C. Implementation

The practical part of this study is the implementation of a platform. As illustrated in Figure 4.2, the application diagram shows the platform internal workflow. This workflow started with an Apache webserver, which handles the communication with both users, and the Twitter Search API, the platform requests messages from Twitter on the JavaScript Object Notation (JSON) format. Then, data is sent to the PHP interpreter, which runs on the Code Igniter Framework. Data is categorized in MCM and forwarded to store in an MySQL database. The researcher used JAVA to develop an OFM, a PDM, and a FCM because of convenience of its libraries. The result from each module is in the text format. Afterwards, WEKA (Waikato Environment for Knowledge Analysis), which is an open-source machine learning software for data mining, is used to construct the opinion model and evaluate the results. Finally, the end result was kept in the MySQL and presented by the SVLM to users via the HTTP protocol.

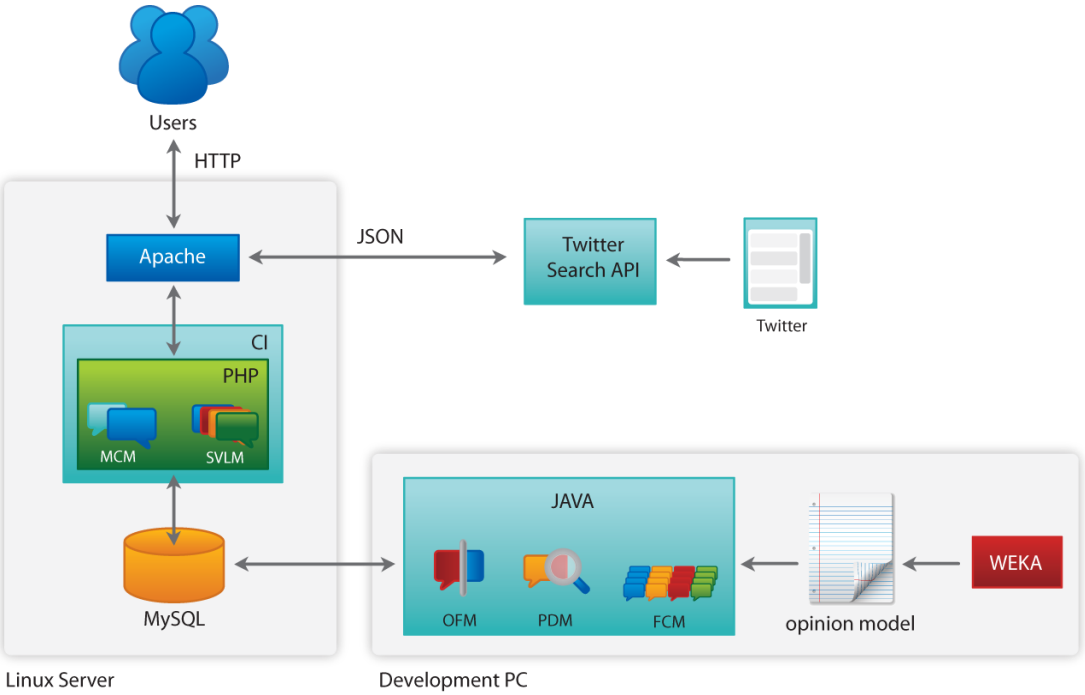

Figure 4.2 Application Diagram

D. Results

To create an OM in the OFM, the researcher applied feature representation and selection techniques, including the information gain (IG) and chi-square (CHI). The researcher compared two classification algorithms, which are Naïve Bayes (NB) and Support Vector Machines (SVM). This evaluation is performed using 10-fold cross validations.

The researcher used data, which consists of 600 tweets, of which 300 are opinions and 300 are others. The 600 tweets were evaluated by WEKA. The total number of extracted term features is equal to 643. To find out the optimal number of term features, the researcher varied the term features from 100 to 600. As shown in Table 4.2, there is not much significant to the classification accuracy by increasing number of features for each method. The result ranges from 82.17 to 84.5%. The approach of the IG on the NB and the CHI on the NB yield almost the same accuracy, which is around 83%, while the IG on SVM and the CHI on SVM provide a little better result of around 84%.

Table 4.2 The performance based on accuracy (%) by varying the number of term features

Number of features	IG + NB	CHI + NB	IG + SVM	CHI + SVM
100	82.83	83.33	82.17	83.00
200	83.83	83.33	83.00	83.00
300	83.00	83.17	84.50	84.00
400	83.00	83.00	84.50	83.83
500	83.00	83.00	84.50	84.00
600	83.00	83.00	83.33	83.33

Next, the classification accuracy comparisons among two algorithms (NB and SVM) on each feature representation, and selection method are shown in Table 4.3. It can be noticed that the accuracy of both the NB and SVM classification for every feature selection methods is over 80%. The highest result comes from the IG feature selection on SVM algorithm, which is higher than the CHI on SVM up to 0.5% and is still higher than the BOW method on SVM up to 2.67%.

Table 4.3 Classification algorithm comparison

Feature selection	Classification algorithm	
	NB	SVM
BOW	83.00	81.83
IG	83.83	84.50
CHI	83.33	84.00

After obtaining the best result with the IG and SVM, the researcher used 10 sets of data, which each contains 10,000 tweets, to validate the classification model. Table 4.4 shows that 19.9% of tweets are classified as opinion tweets. Therefore, the workload of processing is reduced.

Table 4.4 Number of tweets classified as opinion

Sub dataset	Tweets classified as opinion
1	2,156
2	2,036
3	2,111
4	1,817
5	1,756
6	1,784
7	2,025
8	2,285
9	1,725
10	2,208

The experimental results in Table 4.5 show that during the study period iPhone is the brand that people talk about the most, followed by Blackberry and HTC, respectively. However, after the PDM processing, the rank is changed a little to iPhone, Blackberry, and Nexus. The researcher rarely found conversation about Palm and Nokia in the micro-blogs.

Table 4.5 Number and percentage of the twitter posts

	MCM	OFM		PDM	
Blackberry	100,000	4,105	20.63%	548	13.72%
HTC		2,864	14.39%	428	10.71%
iPhone		8,091	40.65%	2,231	55.84%
Motorola		990	4.97%	157	3.93%
Nexus		2,305	11.58%	458	11.46%
Nokia		370	1.86%	48	1.20%
Palm		73	0.37%	16	0.40%
Samsung		1,105	5.55%	109	2.73%
Total	100,000	19,903	100%	3,995	100%

Table 4.6 Number of sentiment tweets for each brand from FCM

	Blackberry		HTC		iPhone		Motorola	
	+	-	+	-	+	-	+	-
Accessories	9	2	3	0	6	14	2	1
Application	150	106	43	20	1071	473	3	3
Camera	66	60	138	25	176	73	36	7
Capacity	2	4	1	1	4	3	0	1
Network	20	21	52	15	104	69	18	12
Power	53	51	65	39	101	123	35	12
Screen	1	0	1	0	4	1	0	0
Sensor	1	0	0	0	0	0	0	0
Size	2	0	3	0	0	0	0	0
System	0	0	21	1	7	1	12	15
	Nexus		Nokia		Palm		Samsung	
	+	-	+	-	+	-	+	-
Accessories	1	3	5	0	1	0	0	2
Application	37	12	2	4	7	1	10	10
Camera	175	139	6	3	2	0	62	7
Capacity	1	0	1	1	0	0	0	1
Network	33	14	17	5	1	0	3	2
Power	16	19	4	0	4	0	4	1
Screen	0	1	0	0	0	0	7	0
Sensor	1	0	0	0	0	0	0	0
Size	0	2	0	0	0	0	0	0
System	4	0	0	0	0	0	0	0

As illustrated in Table 4.6, people prefer talking about the features, Application, Camera, and Power. Surprisingly, the number of complimented tweets outnumbers the number of complaints. For Blackberry, both positive and negative tweets for each feature are almost equally balanced; except the positive tweets for application is more than negative ones. For HTC, the number of positives outperforms the negatives in every feature. For iPhone, the favorite features are Application, Network, and Camera.

However, they have some concern about Accessories. For Motorola, the positive tweets are on Camera and Power. Similar to HTC, Nexus has good perception for every feature except Power. For Nokia, it is strong for their Network. For Palm, there is too little data to say whether it is good or bad. Finally, for Samsung, it is very outstanding for its Camera.

To summarize the result, processes of SVLM will visualize these numbers by three different techniques.

Radar chart – to compare each feature in a brand as illustrated in Figure 4.3. For example, the researcher considered HTC in 10 product aspects. Camera was obviously the best one. Power might be a majorly concerned feature since there are both positive and negative opinions in an interesting portion.

Figure 4.3 The comparisons of features for each brand

Bar chart – to compare each brand for the same feature as illustrated in Figure 4.4. For example, the researcher considered network feature. iPhone was the most talked about one while there were both high positive and high negative amount of opinions. HTC was quite impressive since there were double to three time positive opinions more than negative opinions.

Figure 4.4 The comparisons of brands for each feature

Line chart – to compare selected brand in every features as illustrated in Figure 4.5. For example, the researcher compared Blackberry with Motorola and Nexus.

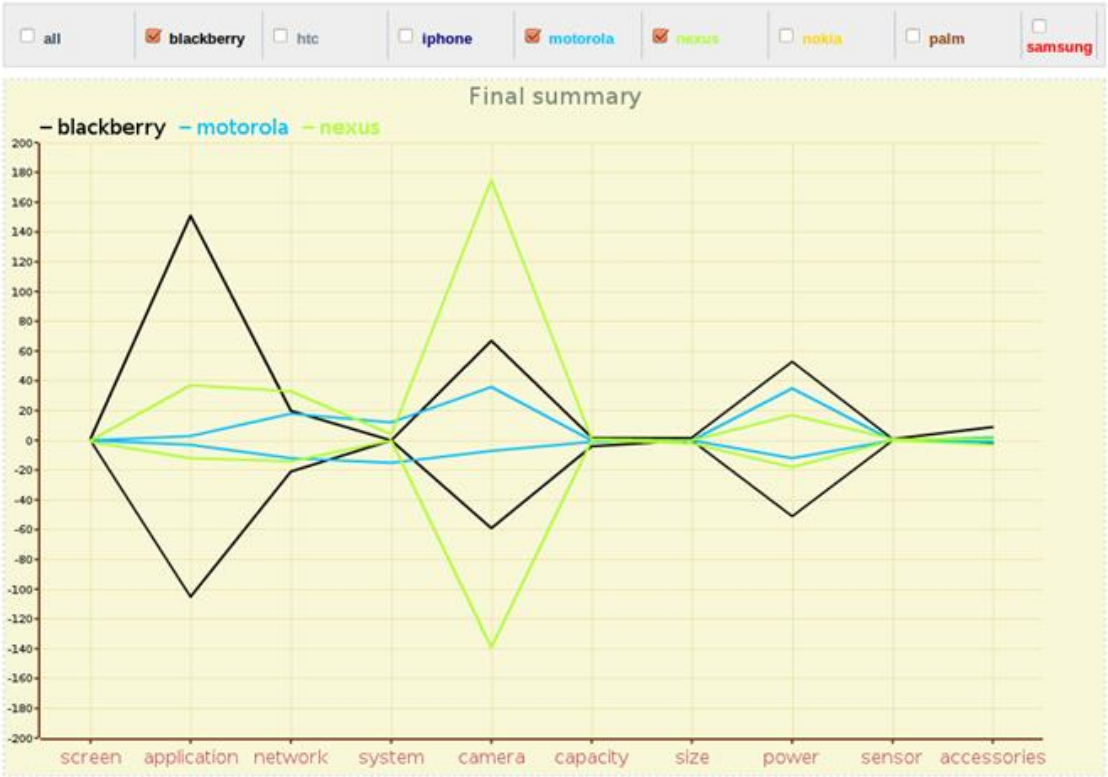

Figure 4.5 The comparisons of brands and features

The final focus group discussion

A. Venue and schedule

The discussion was conducted on 17th February 2012, 13:00-16:00 at 14th floor Chamchuree Square, Technopreneurship and Innovation Management Program, Chulalongkorn University.

B. Participants

Participants composed of nine experts in social media and smartphone industry as follows.

1 Mobile operator

1 Mobile phone Magazine

1 Software development

2 Lecturers from business school and computer science

4 Media agency

C. Result

Objective 1: To evaluate concept of the Micro-blog Sentiment Analysis System prototype

- One participant studied about their customers from corporate database that retrieved from a log file of customers' mobile usage. It is real data showing how many traffics or what applications that users use.
- One participant claimed that he study from twitter by manually read.
- One participant implemented system to capture what happen online; however, it is required human read.
- One participant used many tools before and found problem regarding language and ability to analyze Thai language of each tool.

Table 4.7 Experts opinion on prototype result

	Agree	Disagree	No comment
Blackberry	8	0	1
HTC	7	1	1
iPhone	9	0	0
Motorola	6	1	2
Nexus	6	0	3
Nokia	7	2	0
Palm	4	0	5
Samsung	4	3	1

Regarding to the accuracy of the program, below is the feedback from participants:

- One participant questioned about HTC application
- One participant questioned about Motorola should be more popular than Samsung in the time that collected data. and was curious why so little people talk about Nokia
- One participant mentioned that some messages should be separated out first because it is hired to post from brands; and also some from influencer who are loyal to specific brand. Therefore, users who post is really important to be considered.
- One participant added that the software should protect this kind of spam by filtering users. For example, only people who have follow more than some certain number of people.
- One participant claimed that marketers can use this kind of information for brand positioning using strategy map position.

As illustrated in Table 4.7, experts provided their judgment whether the result of each brand is appropriated or not.

Objective 2: To assess the software prototype acceptance

- One participant suggested the best approach to do sentiment analysis is to combine between a machine and a human coupled with also eliminate noise from a specific user. The Machine should be used to monitor while human help considering which message is significant. Sometimes if user is influencer, weight will gain more confident. Moreover, this is opportunity to launch this type of products in Asia.
- One participant said that ranking is important to show the distinguishing and weighting for significant messages between programs.
- One participant summarized 3 key success factors to concern
    1) the language problem
    2) the habit and behavior
    3) the target audience who use twitter
- One participant suggested changing visitors to buyer by finding idea in communication from the software.
- One participant mentioned that she tried to find a key benefit or a key selling point of the products. This software should be able to find keywords or patterns that suit specific customers

As shown in Table 4.8, it is the summary of scores that participants present for a focus group discussion guide. These composed with three parts, which are the Perceive Usefulness, the Perceive Ease of use, and the Attitude about the software prototype.

Then, the researcher translates the raw score into five-point Likert scale by dividing the result with nine, which calculated by dividing 45 by 5. As a result, the Perceived Usefulness gets 4.19 or “agree” and the Perceived Ease of use gets 3.89 or should be scaled between “neutral” and “agree”. Finally, the Attitude can be transformed to “strongly agree” since every participant answered “yes” to the questionnaire.

Objective 3: Other comments that the experts may have for this prototype

- One participant said he like this software because it is automated. It can reduce many processes for the market research. The result from this software shows the overview of the research result. He does not concern much about the accuracy since if the result is against the intuition, he suggested doing the traditional marketing research in addition. Moreover, he mentioned that this software will benefit for some categories of product only, especially luxury items.
- One participant told that this kind of tools will help both producers and consumers. Moreover, the result from the software should not be older than one month. She also concerned that people who use the micro-blog mostly live in an urban area.
- One participant mentioned that the marketing intelligence information will be useful if the marketer knows exactly who said those messages.
- One participant told that this kind of software will be very useful in the future when more people talk in the micro-blog and the number of messages is too long for human to read.
- One participant required software to be more intelligent by a self-learning keyword.
- It should have predefined variables by the industry
- One participant said that consumers have many problems with the product and wait for brand to solve it. Therefore, brand should take consideration and action to this message.
- One participant suggested providing the package that users can login to use this software by themselves.
- One participant added that Freemium model might be good option.

Table 4.8 Expert acceptance

Perceived Usefulness -> “agree”

To safe time in consumer need investigation	38
To increase effectiveness and precision for understanding consumer need	36
To examine new information that never known before.	39
average	37.67

Perceived Ease of use -> “agree”

Easy to understand	36
Easy for further use in organization	35
Reduce complication in consumer research	34
average	35

Attitude -> “strongly agree”

Interested in use in the future	9/9 participants
Believe that this concept will be able to guide the new product development	9/9 participants
This concept is new in the industry	9/9 participants

CHAPTER V
BUSINESS MODEL INNOVATION

Business requirement

The foregoing chapter will emphasize the opinion mining as an innovative marketing tool for corporate. Currently, marketers have to understand how both current and potential customers are affected by the micro-blog. Since people always gather information from the Internet about products or services they would like to pay, companies must know what they can gain or risk from this feedback on the micro-blog to stay competitive.

However, from the result of Chapter IV, the researcher can notice that providing just a tool is not enough for business needs. In fact, the real demand is the business intelligence. Unfortunately, the Sentiment Analysis software can provide information, but not the knowledge. The lacking piece should be the analysis part or analyst who understands the result and be able to transform the numeric information into the practical knowledge for further deployment.

In the second focus group discussion, the researcher also opened the floor to talk about this losing part. The moderator shoot the question that “Should we provide the service for product repositioning?” The result is listed below:

- One participant provided an example of one cosmetic brand. There are many items under this brand. They did not know before that many online people are looking for one item. However, the social media monitoring tool can provide this information to the marketing department. The cause is that they sell this item in supermarkets but customers mostly look in convenience stores. As a result, they have to relocate the product selling place. This shows that besides finding strong or weak points, the tool can tell details that the company forgets, but other people in the social media saw it.
- One participant added about customer driven marketing strategies and

customer driving marketing strategies. Sometimes people know what they need while sometimes they do not know until others show them. This tool should provide this kind of information for the brand.

- One mentioned that Steve Jobs had told that "Customers can tell you what they 'want'. They know what they want. Unluckily, people cannot say about what they need. After all they do not know what they do not know."
- One participant suggested that this tool should be able to set the target group of speakers and extract keywords that related to the business needs.

Johari Window model

Besides, one related concept is the Johari window model, which is a simple and useful tool for disclosure of self-awareness and mutual understanding between individuals within a group (Chapman, 2003). Like some other behavioral models, the Johari window is based on a four-block area. The standard one shows each quadrant the same size to represent the information in terms of whether the information is recognized or undetermined by self, and whether the information is recognized or undetermined by others. Each quadrant is able be changed in size to reflect the relevant proportions of knowledge they gain.

1. Open/Free area – what is realized by both self and others
2. Blind area – what is unknown by self, but which is known by others
3. Hidden area – what is known by self, but the others do not know
4. Unknown area – what is unfamiliar by self and is also undetermined by others

Information is a key resource using to make decisions. The understanding in each area of the Johari window can help companies to provide products/services that have an edge over competitors. The company will be able to improve the communication and interaction to meet the challenges of today's environment. To move

from area 1 to 2, the feedback satisfaction must be asked. To move from area 1 to 3, the details of the company must be told or exposed more to others. Lastly, to move to area 4, the current situation and methods to fill out the gap must be uncovered.

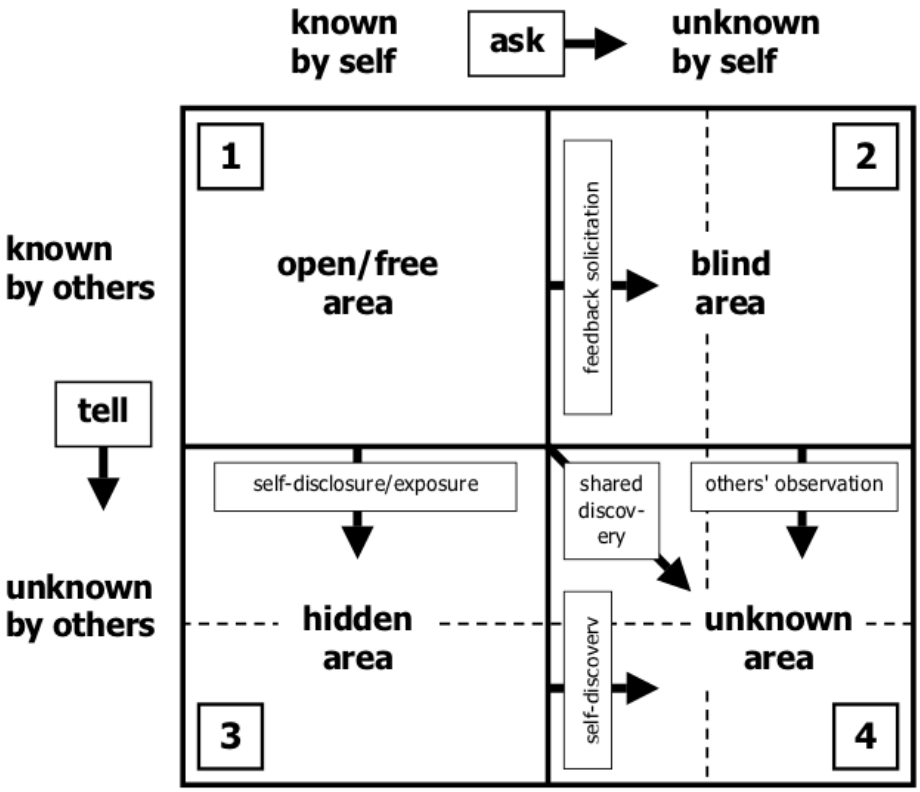

Figure 5.1 Johari Window

For example, this kind of below message may be found in an open/free area.

“I really hate how quicky iPhone battery runs out!!!! But I love how clear the screen is!!”

It can be implied that both Apple and consumers also know that battery should be improved while Apple is also able to identify from consumers that iPhone screen is good enough.

For the blind area, this message should be found.

“Still struggling 2 come 2 terms with the fact that technophobic wife has iPhone. Love my EricssonX1 but it’s not as good on fb or twitter as hers”

It can be implied that facebook and twitter application on EricssonX1 may be

difficult than those on iPhone.

For the hidden area, it is a range that consumer may not realize in some information that the company concerns.

“Being an iPhone developer, awesome, working with iPhone certificates not awesome...”

It can be implied that iPhone developers love their jobs but they may hate the process of the certificate handling, which Apple may have some information why the company have to make it is so complicated. Therefore, if Apple can communicate this information to developers, this issue will be solved in the future.

Lastly, the unknown area is the most difficult to work with. The research demonstrates three messages that lead to one conclusion as follows.

“@lucyhedges I actually just don't use the hardware keyboard much. The touch keyboard is good... though at that rate, get a Nexus One!”

“I think the only bad point about iphone is its keypad, i hate messaging with it.”

“@tri_Heather The wife has an Eris. She likes it. The Nexus One looks pretty cool, but it only has a virtual keyboard. To me that's a bummer.”

These can be inferred that users might have a problem with a keyboard but they do not know what they want while the company also has no clue how to solve this problem. Anyway, these messages may lead to the company to research more in the technology about the keyboard.

Business model canvas

Assumed that the micro-blog research company is established, the researcher will try to make the picture more concrete by embracing the business model innovation to the service. (Osterwalder and Pignerur, 2010) defines that a business model describes the rationale of how an organization creates, delivers, and captures value through nine basic building blocks, call “business model canvas”. As shown in Figure

5.2, the business model canvas can be described as follows.

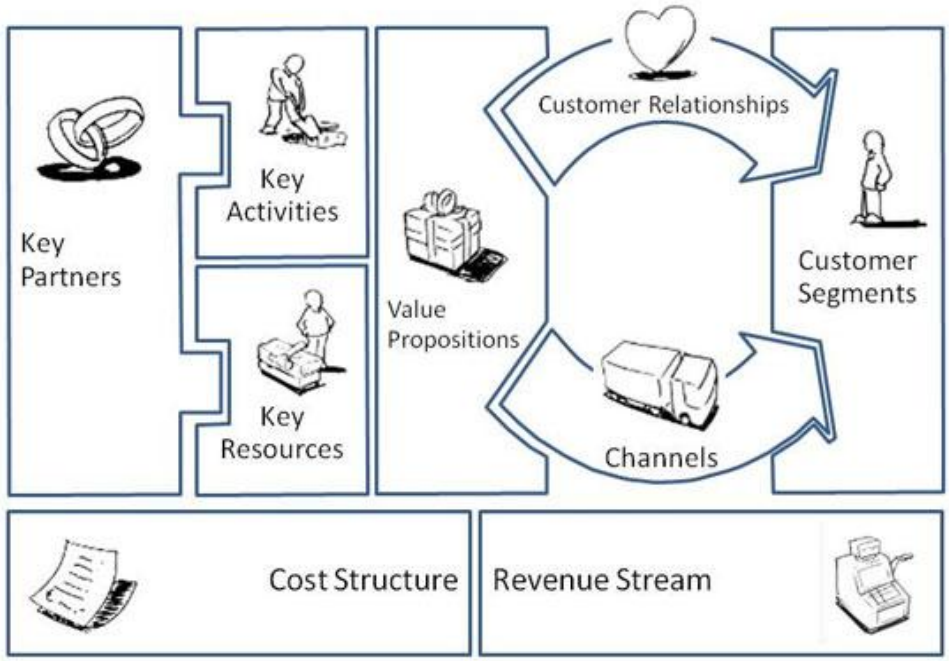

Figure 5.2 The Business model canvas

A. Customer Segments (CS)

It is the different groups of people or organizations that company targets to reach. Established corporate with the short life cycle of the consumer product is the company's first priority customers. Since the market for this kind of services might be separated into three tiers, which are big corporates, agencies, or opening to public, according to the result of an interview. From the researcher's point of view, the company should target a niche market or the middle tier. As a result, the researcher suggests to targets at the "research firms" which are well-known in the industry but have not implemented their own sentiment analysis system yet, such as Advanced Research Group Co., Ltd., Idea360 Co., Ltd., and Millward Brown Firefly Co.,Ltd.

B. Value Propositions (VP)

It is the bundle of products or services that generates value for particular customers. Since the researcher chooses to serve the niche market, the company's value proposition will be customization for a specific industry. Combining with the

concept of Johari windows, the company can provide these kinds of customized services related to the four areas as follows.

1. Open/Free area

This service is to cross check some hypothesis. The clients tell the company about their products/services, and the company will check the responding from the system for what people think about them.

2. Blind area

This service is to find out objects that customers have ignored. The company will try to check around the products/services which features that people talk about but the clients had not paid attention to. Therefore, the clients will be able to improve their products/services.

3. Hidden area

This service is to look for the way to improve the clients' communication. This will be very useful to check the performance of campaign or promotion that the clients invested in. It will increase the return on investment for the clients.

4. Unknown area

This is the most difficult area to work for since the company has no information from the clients and people may talk about it but they do not know what it is. Therefore, the company has to work closely together with the clients to investigate each message. It will be very useful to develop a new product/service in the future.

C. Channels (CH)

It is medium how the company can communicate with the clients to provide a value proposition to them. For owned channels, the company will employ sales force and web sites to direct contacts to the clients. For partner channels, they are indirect and span to agencies. The researcher thinks that the best channel to reach the clients is the direct contact with a good relationship by sale persons.

D. Customer Relationships (CR)

It is the type of relationships that the company will try to establish with specific

clients. The company will create its connections based on the human interaction by dedicated personal assistances. Since the nature of this service requires the company to work closely with clients, the company will need the most sincerely associated and friendly developed connections over a long period of time.

E. Revenue Streams (R$)

It is the income that the company creates from each client segment. The company's revenue comes from the usage fee based on the product feature dependent price, which the price determined on the quality of the value propositions. As mentioned in Value Proposition part, the researcher separated the service to four levels related to the four areas of Johari windows. The company should charge the clients based on the answer that they would like to know and the amount of data has to be analyzed. However, the company might provide the free trial service for the first-time clients and also publish general report for the benefit of the whole industry.

F. Key Resources (KR)

It is one of the most important factors required to make a business model operate. This service requires intellectual resource and human resource. Though, software is protected by copyrights, the idea of develop this kind of software is not protected. The researcher has to continuously improve and add some more features all the time. Therefore, staffs are the most significant factor for the key resource. Two main functions are software development and sales department. While the company tries to recruit good programmers to join the team, the company has to have a lot of convincing sale people to sell the service.

G. Key Activities (KA)

It is the most important thing that the company must perform to create its business model process. For this service, key activities start from developing the software to sell the service and supporting customers. The critical point may be the first step to find out and delivery requests from clients.

H. Key Partnerships (KP)

It is the network partners or suppliers that push the business model works. There

are four main types of partnerships: (1) strategic alliances between non-competitors (2) Competition, strategic alliances between non-competitors (3) Joint ventures to develop new businesses (4) Buyer-supplier relationships to assure reliable suppliers. In the beginning of the business, the company should maintain the buyer-supplier relationship with research firms or agencies that are contacted. Moreover, the most important partners are media companies since they will be the one to promote the service indirectly to potential customers.

I. Cost (C$)

It is all expenses incurred to run this business model. Since this service is not the cost-driven, the company has to focus on the value creation. A premium value proposition and a high degree of customization is the drive of this kind of businesses. The payment of the company is the staff salary and the technical expense to maintain the server.

As a result, the researcher has put these values into table as shown in Figure 5.3

<table>
<tr><td rowspan="2">KP

Media</td><td>KA

Software development,
Service selling,
Customer support</td><td rowspan="2">VP

Customize service for specific industry</td><td>CR

Dedicated personal assistances</td><td rowspan="2">CS

Research Firms</td></tr>
<tr><td>KR

People

Lexicon</td><td>CH

Sales people

Website</td></tr>
<tr><td colspan="2">C$

Staff / Sever</td><td colspan="3">R$

Usage fee</td></tr>
</table>

Figure 5.3 MSAS Business Model canvas

Financial plan

As a part of commercialization, a financial plan is required to provide a set of the cash flow projection. Starting from cost calculation, the researcher put all related costs through three years of study period in Table 5.1. The cost composed of the researcher wage, the advisor wage, the server rental fee, and other R&D expenses, such as programing and focus group discussion.

Table 5.1 Cost of system development

	Months	Unit	Cost/month	Total
Researcher	36	1	20,000	720,000
Advisors	18	2	5,000	180,000
Server		1		30,000
R&D expenses		1		100,000
				1,030,000

Next, the researcher estimates both revenues and expenses to form up the financial forecast for four years as illustrated in Table 5.2. The service price is set up at THB 25,000 per year per client. The researcher expects 100 accumulated client per year. For the expense, it should be composed of three main portions which are the staff salary, the server rental fee, and the management fee.

The company should have at least five staffs, which are rated around THB 25,000 per month, to support 100 clients. The rental fee for a server is approximately THB 100,000 per year. Lastly, the management fee, such as office rental fee, utilities, and other operation costs is estimated around THB 500,000 for the whole year. Therefore, the expense for the first year is equal THB 2,100,000.

The revenue is expected to grow double in the second year, triple in the third year, and four times in the fourth year. Then, the number of staff, and the server will be

increased at the same rate as revenue while the management fee might be increase 50% per year. At the end of the fourth year, the net profile is approximately at THB 3,487,500.

Table 5.2 Four years financial forecast for MSAS service

	Year 1	Year 2	Year 3	Year 4
Revenue	2,500,000	5,000,000	7,500,000	10,000,000
Expense				
Staff	1,500,000	3,000,000	4,500,000	6,000,000
Server	100,000	200,000	300,000	400,000
Management	500,000	750,000	1,125,000	1,687,500
Sum	2,100,000	3,950,000	5,925,000	8,087,500
Net	400,000	1,050,000	1,575,000	1,912,500
Balance	400,000	1,450,000	2,625,000	3,487,500

CHAPTER VI
EPILOGUE

Discussion and implications

In this chapter the empirical findings are analyzed. There are both qualitative and quantitative in the study. The analysis of the qualitative data is remarkably different from the statistic method, as the data is not presented in a numerical form. When analyzing interviews and focus group discussions, it is difficult to set numerical descriptions to the summary. Conversely, in the quantitative part, the researcher obtained straight forward number from the experiment. These activities enlarge the knowledge of both theory and empirical practices.

A. The initial focus group discussion

Objective 1: To understand the use of the micro-blog in an organization

Currently, every organization realize in the impact of the micro-blog to their works. Some companies might concern and have restricted policy for using the micro-blog during office hours. The different perspective between the traditional management style and the contemporary management style still existed. For the traditional one, executives focus employees to work exclusively on their jobs. For the contemporary style, executives allow employees to gather information from many available sources as possible. However, most companies still use the micro-blog for announcing their information only. Just a few companies use the micro-blog for listening events happened in the Internet society.

Objective 2: To determine the need of the micro-blog for the market research

After the discussion, all participants were interested in using the micro-blog as a source for listening voice of consumers. However, they had mainly three concerns as follow:

1. Amount of usable data – It is not guarantee about the quantity of information that collected from the micro-blog. It may be overwhelming or deficient for using in the research.

2. Language – since the written language in the Internet is informal, marketers can expect a lot of slang and abbreviation that sometimes is very difficult to answer even by human.

3. Influencing – it is very easy to find viral phenomenon in the micro-blog, especially when some influencers speak and other believe without deliberation. This may affect the results of the crowd opinions in just a short period of time.

Objective 3: To gather requirement for the software prototype development

There are many opportunities in the software development regarding OM on the micro-blog. The companies can use in many activities, such as R&D, CRM, and Marketing. However, there is concern about catchment and sampling. If the software can bring up a statistic method, it will yield high validity and confidentiality.

Objective 4: To find business opportunity for the market research service on the micro-blog

Sine this software is quite difficult to develop in-house, every company was interested in using the service. The price depends on for what the result can be applied. However, there was a suggestion that the provider who provides this service must have high creditability. However, not everyone is able to offer this kind of service for commercial use. One more interesting suggestion is that the

provider should supply aggregate information for the benefit of the whole industry.

B. In-depth interview

Below are issues that the researcher aggregates from the idea of five experts for the direction of the software development

Main idea:

- Since each industry has different characters and environment, the software should work well depended on the industry.
- This software should be able to use to confirm an idea or some hypothesis.
- Tool should be able to determine the strongest point to communicate back to consumers.
- Users could use this software complimentary with the tradition marketing research.
- Tool must be traceable with time and be able to predict trend in the future
- Tool must be integrated with the legacy system to facilitate employees for their routine jobs.
- Tool should have feature to analyze speaker coupled with the message and time. It will be nice if the tool can find out the person who is the influencer in each industry. Corporate would like to recognize the first person who ignites the issue.

Presentation part:

- Tool should be able to show top ten keywords for both positive and negative sides.
- Final results should be presented in a graph report.

Processing part:

- Time for processing must be very short or ideally it should be the real-time analysis

Spam handling part:

- Tool must be able to distinguish between spam and the real opinions
- Tool should be able to check the consistency of speaking

C. Software experiment

In this experiment, the system was performed on the combination of both the lexicon-based and the machine learning based approaches. Like Pang's work (Pang et al., 2002), the researcher used the machine learning approach in the Opinion Filtering Module (OFM) to obtain high recall while SentiWordNet which based on the lexicon-based approach, using in the Polarity Detecting Module (PDM), can deliver high precision (Baccianella et al., 2010).

Table 6.1 Example of other tweets which are misclassified as opinions

No.	Example
1	Paris Hilton: Love My New Billboards up in Times Square NYC! Huge! :) http://twitpic.com/27d3qj : @ParisHilton love wot u do wid ur BB :)
2	Had a good session @LesMills with @kirstwgtn2010 tonight!! Except I was a tard and dropped the iPhone... Seems to be alive but a lil bruised
3	Slice Home | Handheld Pda And Accessories: fr HTC TOUCH HD T8282 CELL PHONE FAST WALL HOME CHARGER Paypal US $7.89... http://bit.ly/9HJiYE
4	Update: Motorola Cliq With T-Mobile Activation Plan http://www.uugame.tk/android/motorola-cliq-with-t-mobile-activation-plan
5	@borstal_boy good morning BB. have you been to a doctor?

Table 6.1 and 6.2 show some challenging examples for classifying

opinions and others. Table 6.1 provides some example of other tweets which are misclassified as opinions while Table 6.2 presents some examples of opinionated tweets which are misclassified as others. The difficulties are caused by many reasons, including the language semantic and the need of world knowledge. From Table 6.1, post #1 contains a narrative sentence which is classified as an opinion, i.e., “love wot u do wid ur BB”. Using the simple term feature would simply recognize “love” as a positive polar word; therefore the classification model could be confused in learning such case.

From Table 6.2, post #6 is a negative opinionated sentence, i.e., “I cannot type some of the alphabet on software keyboard and more trouble.” However, this tweet is misclassified due to the missing of explicit negative polar words. Post #7 is also difficult to identify since it contains a comparative sentence of one product, i.e., Blackberry, to others, i.e., iPhone and Android.

Table 6.2 Example of opinionated tweets which are misclassified as others

No.	Example
6	HTC EVO 4G may have touch screen related hidden problem. I cannot type some of the alphabet on software keyboard and more trouble.
7	Playing with my son's Blackberry Curve Easy to see why iPhone, Android killing them. #pathetic
8	@PuRpLe_KisSes89 i have the HTC Touch Pro 2 but yeah this app is unreliable so i end going on mapquest anyways
9	@sarat touch should be like ipod. nokia n samsung is not so cool!
10	@ejcarey Both are great phones. Would be nice to see another manufacturer step up to Apple like HTC has. Push Apple a bit more.

Although there have been some previous researches in the sentiment analysis on the micro-blog, no research has emphasized on the utilization of the end result. The result from Twitter Sentiment (Go et al., 2009) provides only

overview of consumers' sentiments on a product, but it cannot specify the sentiment on any exact feature.

However, the result from the MSAS may not yield direct feedback for the product feature itself. For example, the result of iPhone shows that Apple might have some problems with its accessories, but the fact is most accessories are not manufactured by Apple. Apple allows third-party companies to build accessories for their iPhone. Therefore, the results from this system must have to comply with the real world situation first.

D. The final focus group discussion

Objective 1: To evaluate concept of the Micro-blog Sentiment Analysis System prototype

The concept of OM on the micro-blog is admitted in this participant group since there are many international related products available in the market though it is not popular in Thailand yet. Some participants used this kind of tools before and realize in the limitation of the current software. The feedback of prototype that is the outcome from the expert evaluation can be summarized below:

1. Agree

This will be happened when people can gain enough data for the system. iPhone is obvious case because there are tremendous amount of data. Everybody agrees with the result of iPhone. Next, Blackberry, HTC, and Motorola should be classified in this group. Although there was a participant disagreed with the result, most participants did not have any doubt.

2. Disagree

The researcher notices that participants may not believe the result when the system could not retrieve enough data from the micro-blog. The researcher

put Nokia and Samsung in this group. There were 2 participants doubted in Nokia while there were 3 participants doubted in Samsung. Samsung is an exceptional case since it is very popular now but there was too little people talking about it 2 years ago when the system collected data.

3. No comment

Since some participants were not familiar with Nexus, which official available in the U.S., and Palm, which is not popular in Thailand, it is hard to describe some notion on them.

In summary, the researcher can see patterns that the result will be more accepted when enough data is collected. Conversely, the result will be more questionable when data is not able to be adequately collected. Therefore, the number of data is the most significant factor for reliability of this software.

Objective 2: To assess the software prototype acceptance.

As illustrated in Figure 6.1, the researcher separates questions into three main parts. First, perceived usefulness is rated 37.67 from 45. Second, perceived ease of use is rated 35 from 45. Finally, all participants have a good attitude to this prototype. Every participant was interested in using this service in the future. They accepted that this service would be able to guide the new product development in their industries, and consented that this service is new in the industry.

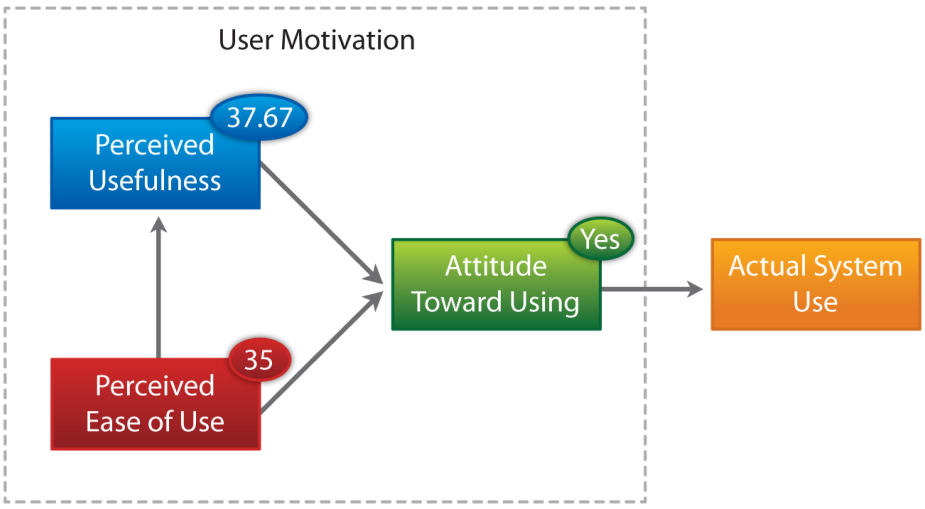

Figure 6.1 The result of the MSAS evaluation by applying the TAM concept

Conclusion

In the software experiment part, a solution based on the standard well-known text classification technique to automatically filter opinion tweets from other types of tweets is presented to enhance the performance of sentiment analysis on Twitter. The proposed intention classification consists of three steps: (1) the text processing (2) the feature processing and (3) the classification model construction. From this experiment, the Support Vector Machines (SVM) algorithm with the Information Gain (IG) feature selection yielded the best performance with the accuracy equal to 84.5%. Then, the researcher validated the proposed model and found that 19.9% of tweets are classified as opinion posts.

The researcher reported an exploratory study using the Micro-blog Sentiment Analysis System (MSAS) to discover the consumer insight. The work reconfirms that the sentiment analysis on the micro-blog, especially Twitter, can provide supportive information for producers in the smartphone industry to process some decision about their next generation products. The MSAS can gather information regarding the product feature reviews without disturbing consumers, and the result is acceptable by experts in the field. According to Technology Adoption Life Cycle, the researcher considers that MSAS is suitable for studying consumer opinions in 2 periods: first, it will be useful in the early adoption period for adjusting messages to communicate to the public, second, if

researcher collect data in the late majority period, the result will be useful for a new product development for the next product generation.

Consequently, it can be concluded that the sentiment analysis on the micro-blog is very useful tool for the consumer research, especially in the industries that customers spend their time on the social media. Many departments in corporate can adopt this sentiment analysis tool to support their tasks. Mainly, PR and marketing department can learn from consumer messages how they react to the product/service of the company. R&D department can use this information to improve or develop new product. CRM department can plan their activities based on what customer want. However, the less data the system can collect from social media, the more errors the system will suffer from the analyzed result. Comparing with other products in the market, the MSAS is better in the way that it can analyze sentiment into the feature level of the products, which will be more valuable for marketers.

Future works

For future works, the researcher plan to improve the accuracy of the Opinion Filtering Module (OFM) by three following ways. The first is by adopting the active learning technique to automatically increase the training corpus size. Using more training posts for constructing a classification model would generally help increasing the classification model effectiveness. The second approach is to improve the feature processing technique by (1) performing posts normalization, i.e., transforming chat language into the normal written language, and (2) representing features with syntactic units, such as phrases with part-of-speech (POS) tags, instead of terms. For example, the phrase “cannot type” carries a negative sense for the smartphone even though it does not contain an explicit negative polar word. Therefore, include syntactic units as additional features would also likely to increase the overall classification performance. The third approach is to study the data that classified non-opinion from the OFM. Since this module separates around 80% of non-opinion posts from the whole messages, there will be potential to have fault results in this data. Users may input this data to the OFM

again to study the result of these non-opinion posts whether it is correct or wrong.

The future work is to increase the accuracy of the analyzed results. More research on algorithms will be performed to improve the performance and accuracy of the system. The researcher would also study some solutions to handling some interesting and challenging issues, such as monitoring the opinions and sentiment from each single user over time. Other issues are to improve the algorithm in order to (1) handle tweets containing multiple features and multiple polarities in the same post, and (2) to handle multiple negations in the same post.

Finally, to commercialize this service, many more features should be implemented, especially the features related to investigate users. The researcher needs integrating the knowledge about the Social Networking Analysis to the software, and expanding from the micro-blog to other sources available on the Internet, such as forum and e-mail.

REFERENCES

Agarwal, A., Xie, B., Vovsha, I., Rambow, O., and Passonneau, R. Sentiment analysis of Twitter data. Proceedings of the Workshop on Languages in Social Media (LSM 2011), pp.30-38, Association for Computational Linguistics, 2011.

Baccianella, S., Esuli, A., and Sebastiani, F. SENTIWORDNET 3.0: An Enhanced Lexical Resource for Sentiment Analysis and Opinion Mining. Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC 2010), European Language Resources Association [Online]. 2010. Available from : http://www.lrec-conf.org/proceedings/lrec2010/pdf/769_Paper.pdf [2011, June]

Bansal, M., Cardi, C., and Lee, L. The power of negative thinking: Exploring label disagreement in the min-cut classification framework. Proceedings of the International Conference in Computational Linguistics (COLING), pp. 13–16, [Online]. 2008. Available from : http://www.cs.cornell.edu/home/llee/papers/disagree.pdf [2011, January]

Bobo, LD., and Dawson, MC. A change has come: Race, politics, and the path to the Obama presidency. Du Bois Review: Social Science Research on Race 6(1), pp.1-147, [Online]. 2009. Available from : http://scholar.harvard.edu/sites/scholar.iq.harvard.edu/files/bobo/files/change.pdf [2011, January]

Boyd, D., Golder, S., and Lotan, G. Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter. Proceedings of the 43rd Hawaii International Conference on Social Systems (HICSS), pp. 1-10, Institute of Electrical and Electronics Engineers, 2010.

Cashmore, P. Facebook Founder on Privacy: Public Is the New "Social Norm". [Online]. 2010. Available from : http://mashable.com/2010/01/10/facebook-founder-on-privacy [2010, January]

Chapman, A. Johari window model diagram. [Online]. 2002 Available from : http://www.businessballs.com/johariwindowmodel.htm [2011, December]

Cheng, A., Evans, M., and Singh, M. Inside Twitter: An in-depth look inside the Twitter world. [Online]. 2009. Sysomos Inc., Available from : http://www.sysomos.com/docs/Inside-Twitter-BySysomos.pdf [2010, April]

Constantinides, E., and Fountain, S. Web 2.0: Conceptual foundations and Marketing Issues. Journal of Direct, Data and Digital Marketing Practice Vol 9, nr 3, (2008): 231-244.

Dave, K., Lawrence, S., and Pennock, D.M. Mining the Peanut Gallery: Opinion Extraction and Semantic Classification of Product Reviews. Proceedings of the 12th international conference on World Wide Web (WWW '03), pp. 519-528, Association for Computing Machinery, 2003.

Davis, F.D. A technology acceptance model for empirically testing new end-user information systems: Theory and results. Doctoral dissertation, Sloan School of Management, Massachusetts Institute of Technology, 1986.

D'Monte, L. Swine flu's tweet tweet causes online flutter. [Online]. 2009. Available from : http://www.business-standard.com/india/news/swine-flu%5Cs-tweet-tweet-causes-online-flutter/356604/ [2011, April]

Eckerson, W. Understanding Business Intelligence. What Works Vol 16. Seattle, WA: Data Warehousing Institute, 2003.

Esuli, A., and Sebastiani, F. SentiWordNet: A Publicly Available Lexical Resource for Opinion Mining. Proceedings of the 5th Conference on Language Resources and Evaluation (LREC 2006), pp.417-422, European Language Resources Association, 2006.

Esuli, A., and Sebastiani, F. PageRanking WordNet Synsets: An Application to Opinion Mining. Proceedings of the 45th Annual Meeting of the Association of Computational Linguistics (ACL '07), pp.424-431, Prague, Czech Republic: Association for Computational Linguistics Society, 2007.

Esuli, A., and Sebastiani, F. Machines that learn how to code open-ended survey data. International Journal of Market Research, Vol. 52 issue 6 (2010) : 775 – 800.

Esuli, A., and Sebastiani, F. Sentiment quantification. IEEE Intelligent Systems Vol. 25 Issue 4 Special issue: AI and Opinion Mining, Part 2. (2010) : 72 - 79.

Evelson, B., McNabb, K., Karel, R., and Barnett, J. It's time to reinvent your BI strategy. Intelligent Enterprise, 2007.

Ferreira, A. Android OS changes smartphone life cycle. [Online]. 2011. Available from http://www.theusdvista.com/business/android-os-changes-smartphone-life-cycle-1.2000033 [2011, February]

Gamon, M., Gao, J., Brockett, C., and Klementiev, R. Using contextual speller techniques and language modeling for ESL error correction. Proceedings of the Third International Joint Conference on Natural Language Processing (IJCNLP'08), pp. 449-456, Asia Federation of Natural Language Processing, 2008.

Glance, N., Hurst, M., Nigam, K., Siegler, M., Stockton, R., and Tomokiyo, T. Deriving marketing intelligence from online discussion. Proceedings of the 11th ACM SIGKDD international conference on Knowledge discovery in data mining (KDD '05), pp. 419-428, Association for Computing Machinery, 2005.

Go, A., Bhayani, R., and Huang, L. Twitter sentiment classification using distant supervision. Technical report, Stanford Digital Library Technologies Project, pp.1-6, [Online]. 2009. Available from : http://www.stanford.edu/~alecmgo/papers/TwitterDistantSupervision09.pdf [2011, March]

Hu, M., and Liu, B. Mining and summarizing customer reviews. Proceedings of the 10th ACM SIGKDD international conference on Knowledge discovery and data mining (KDD '04), pp. 168-177, Association for Computing Machinery, 2004.

Jansen, B., Zhang, M., Sobel, K., and Chowdury, A. Micro-blogging as online word of mouth branding. Proceedings of the 27th international conference extended abstracts on Human factors in computing systems (CHI EA '09), pp. 3859-3864, Association for Computing Machinery, 2009.

Java, A., Song, X., Finin, T. and Tseng B. Why we twitter: understanding microblogging usage and communities. Proceedings of the 9th WebKDD and 1st SNA-KDD 2007 workshop on Web mining and social network analysis (WebKDD/SNA-KDD '07), pp.56-65, Association for Computing Machinery, 2007.

Kazeniac, A. Social networks: Facebook takes over top spot, Twitter climbs. [Online]. 2009. Available from : http://blog.compete.com/2009/02/09/facebook-myspace-twitter-social-network/ [2011, February]

Kouloumpis, E., Wilson, T., and Moore, J. Twitter Sentiment Analysis: The Good the Bad and the OMG!. Proceedings of the 5th International AAAI Conference on Weblogs and Social Media (ICWSM-11) [Online]. 2011. Available from : http://www.aaai.org/ocs/index.php/ICWSM/ICWSM11/paper/view/2857 [2011, December]

Kim, S. and E. Hovy. Extracting opinions, opinion holders, and topics expressed in online news media text. Proceedings of the Workshop on Sentiment and Subjectivity in Text (SST '06), pp. 1-8, Association for Computational Linguistics, 2006.

Lake, T. Twitter Sentiment Analysis. Western Michigan University, [Online]. 2011. Available from : http://cs.wmich.edu/~tllake/fileshare/tllake.twittersentiment.pdf [2011, December]

Lindlof, T. R., and Taylor, B. C. Qualitative Communication Research Methods, 2nd Edition. Thousand Oaks, CA: Sage, 2002.

Liu, B., Hu, M., and Cheng, J. Opinion observer: analyzing and comparing opinions on the Web. Proceedings of the 14th international conference on World Wide Web (WWW '05), pp. 342-351, Association for Computing Machinery, 2005.

Liu, B. Opinion Mining and Sentiment Analysis. Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data (Data-Centric Systems and Applications), pp. 459-526, Springer, 2011.

Marwick, A. E., and Boyd, D. I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. New Media & Society Vol. 13 (February 2011), 114-133.

Negash, S. Business Intelligence. Communications of the Association for Information Systems Vol. 13, Article 15. [Online]. 2004. Available from : http://aisel.aisnet.org/cais/vol13/iss1/15 [2011, March]

Osterwalder, A., and Pignerur, Y. Business Model Generation: A Handbook for Visionaries, Game Changers, and Challengers. John Wiley & Sons, 2010.

Pear Analytics. Twitter Study – August 2009. [Online] 2009. Available from : http://www.pearanalytics.com/blog/wp-content/uploads/2010/05/Twitter-Study-August-2009.pdf [2010, December]

Pang, B., Lee, L., and Vaithyanathan, S. Thumbs up?: sentiment classification using machine learning techniques. Proceedings of the ACL-02 conference on Empirical methods in natural language processing (EMNLP '02), pp. 79-86, Association for Computational Linguistics, 2002.

Pang, B., and Lee L. A sentimental education: Sentiment analysis using subjectivity summarization based on minimum cuts. Proceedings of the 42nd Annual Meeting on Association for Computational Linguistics (ACL '04), pp. 271-278, Association for Computational Linguistics Stroudsburg, 2004.

Pang, B., and Lee, L. Opinion Mining and Sentiment Analysis. Journal Foundations and Trends in Information Retrieval Vol 2 issue 1-2 (January 2008): 1-135.

Popescu, A. and Etzioni, O. Extracting product features and opinions from reviews. Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing (HLT '05), pp. 339-346, Association for Computational Linguistics, 2005.

Stone, M., Bond, A., and Foss, B. CONSUMER INSIGHT: How to use data and market research to get closer to your customer, pp.2-15, Kogan Page, London, UK, 2008.

Terdiman, D. To Twitter or Dodgeball at SXSW? Are you on Twitter? If you're at SXSW, it's almost a requirement. [Online]. 2007. Available from http://news.cnet.com/8301-17939_109-9696264-2.html [2011, March]

Thomas, M., Pang, B., and Lee, L. Get out the vote: determining support or opposition from congressional floor-debate transcripts. Proceeding EMNLP '06 Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing, pp. 327-335, Association for Computational Linguistics, 2006.

Turney, P. Thumbs up or thumbs down?: semantic orientation applied to unsupervised classification of reviews. Proceedings of the 40th Annual Meeting on Association for Computational Linguistics (ACL '02), pp. 417-424, Association for Computational Linguistics, 2002.

Turney, P. Similarity of Semantic Relations. Journal Computational Linguistics, Vol. 23 Issue 3 (September 2006), 379-416.

Twitter blog. Your world, more connected. [Online]. 2011. Available from : http://blog.twitter.com/2011/08/your-world-more-connected.html [2011, August]

Twitter website. Using the Twitter Search API. [Online]. 2011. Available from : https://dev.twitter.com/docs/using-search [2011, October]

Wilas Chamlertwat, Pattarasinee Bhattarakosol, and Choochart Haruechaiyasak. Improving Sentiment Analysis on Twitter with Intention Classification. Proceedings of 2011 3rd International Conference on Computer Engineering and Applications (ICCEA 2011), pp.254-258, 2011.

Wilson, T., Wiebe, J., and Hoffmann, P. Recognizing contextual polarity in phrase-level sentiment analysis. Proceedings of the conference on Human Language Technology and Empirical Methods in Natural Language Processing (HLT '05), pp.347-354, 2005.

Zabin, J., and Jefferies, A. 2008. Social Media Monitoring and Analysis: Generating Consumer Insights from Online Conversation. pp.5-9: Aberdeen Group Benchmark Report, 2008.

Zhai, Z., Liu, B., Xu, H., and Jia, P. Clustering product features for opinion mining. Proceedings of the 4th ACM international conference on Web search and data mining (WSDM '11), pp.347-354, Association for Computing Machinery, 2011.

Zhai, Z., Liu, B., Wang, J., Xu, H., and Jia, P. Product Feature Grouping for Opinion Mining Using Soft-Constraints and EM. Journal of IEEE Intelligent Systems Vol. PP Issue 99 (April 2011).

Zhai, Z., Xu, H., Kang, B., and Jia, P. Exploiting Effective Features for Chinese Sentiment Classification. Journal of Expert Systems with Application. Vol 38 Issue 8 (August 2011): 9139-9146.

Zhao, D. and Rosson, M.B. How and why people Twitter: the role that micro-blogging plays in informal communication at work. Proceedings of the ACM 2009 international conference on Supporting group work (GROUP '09), pp. 243-252, Association for Computing Machinery, 2009.

Zhang, L., Ghosh, R., Dekhil, M., Hsu, M., and Liu, B. Combining Lexicon-based and Learning-based Methods for Twitter Sentiment Analysis. HP Laboratories, HPL-2011-89 [Online]. 2011. Available from : http://www.hpl.hp.com/techreports/2011/HPL-2011-89.html [2011, December]

Zhe, X. A Sentiment Analysis Model Integrating Multiple Algorithms and Diverse Features. Master's Thesis, Computer Science and Engineering, Ohio State University [Online]. 2010. Available from http://etd.ohiolink.edu/send-pdf.cgi/Xu%20Zhe.pdf?osu1275400109 [2011, December]

APPENDICES

APPENDIX A

กำหนดการสนทนากลุ่ม (Focus group discussion)
หัวข้อ “Innovative Marketing Tool by Applying Opinion Mining on Micro-blogs”
โดย นายวิลาส ฉ่ำเลิศวัฒน์
ในความดูแลที่ปรึกษาโดย ผศ.ดร.ภัทรสินี ภัทรโกศล และ ดร.ทิพภากร รังคสิริ
วันที่ 23 มิถุนายน 2553 เวลา 13:00 – 16:00 น.
ณ. ห้องบรรยาย 4 หลักสูตรธุรกิจเทคโนโลยีและการจัดการนวัตกรรม จุฬาลงกรณ์มหาวิทยาลัย
ชั้น 14 อาคารจัตุรัสจามจุรี

กำหนดการ

13:00 – 13:15 ลงทะเบียน

13:15 – 13:20 กล่าวต้อนรับ

แนะนำผู้ควบคุมสนทนากลุ่ม ศาสตราจารย์ กิตติคุณ ดร.อัจฉรา จันทร์ฉาย

แนะนำผู้เข้าร่วมสนทนากลุ่ม อันประกอบด้วย

ดร.พงศกร สายเพ็ชร์ (technology - software)
กรรมการผู้จัดการ บริษัท เอเทรียม เทคโนโลยี จำกัด

คุณกิตติพงศ์ กิตติภัสสร (technology - mobile phone)
ผู้อำนวยการฝ่ายบริหารผลิตภัณฑ์ บมจ. ทีดับบลิวแซด คอร์ปอเรชั่น

คุณภาวุธ พงษ์วิทยาภานุ (technology - Internet)
กรรมการผู้จัดการ บริษัท ตลาดดอทคอม จำกัด

คุณชนินทร์ วิศาลบูชนีย์ (advertising agency)
Executive Creative Director TBWA (THAILAND)

คุณพงศ์ ศกุนตนาค (business - business)
กรรมการผู้จัดการ บริษัท ออฟฟิซ คลับ (ไทย) จำกัด

ดร.ณัฐกิตติ์ ตั้งพูลสินธนา (business - consumer)
ผู้ช่วยกรรมการผู้จัดการใหญ่ ฝ่ายการตลาด บมจ. เซ็นทรัลพัฒนา

คุณประเสริฐ ศรีอุฬารพงศ์ (business - consumer)
กรรมการบริหาร บริษัท สยามรีเทล ดีเวลล็อปเม้นท์ จำกัด

13:20 – 13:30 แนะนำวัตถุประสงค์การสนทนากลุ่ม พร้อมนำเสนอภาพรวมของงานวิจัย

13:30 – 14:00 คำถามที่ 1 เพื่อทำความเข้าใจการใช้งาน micro-blog ในองค์กร

14:00 – 14:40 คำถามที่ 2 เพื่อทำความเข้าใจการใช้งาน micro-blog กับการวิจัยตลาด

14:40 – 15:30 คำถามที่ 3 เพื่อหาความต้องการที่ชัดเจนไปใช้ในการพัฒนาซอฟต์แวร์ต้นแบบ

15:30 – 15:50 คำถามที่ 4 เพื่อประเมินความเป็นไปได้ทางธุรกิจของบริการวิจัยการตลาดบน micro-blog

15:50 – 16:00 สรุปการอภิปราย ซักถามข้อเสนอแนะทั่วไป
กล่าวขอบคุณ พร้อมมอบของที่ระลึก

แบบสัมภาษณ์และสนทนากลุ่ม (Focus group discussion)

หัวข้อ “Innovative Marketing Tool by Applying Opinion Mining on Micro-blogs”

วัตถุประสงค์

1. เพื่อทำความเข้าใจการใช้งาน micro-blog ในองค์กร
2. เพื่อทำความเข้าใจการใช้งาน micro-blog กับการวิจัยตลาด
3. เพื่อหาความต้องการที่ชัดเจนไปใช้ในการพัฒนาซอฟต์แวร์ต้นแบบ
4. เพื่อประเมินความเป็นไปได้ทางธุรกิจของบริการวิจัยการตลาดบน micro-blog

วัตถุประสงค์ 1 เพื่อทำความเข้าใจการใช้งาน micro-blog ในองค์กร (30 นาที)

1. ท่านเคยใช้ micro-blog หรือไม่

   □ เคย □ ไม่เคย

   หาก เคย ใช้มาก่อน เพื่อวัตถุประสงค์ใด

   □ สนทนา □ ติดตามข่าวสาร □ ใช้ในธุรกิจ □ แค่ทดลองเล่น

   □ อื่น ๆ โปรดระบุ _____

2. ท่านคิดว่า micro-blog มีประโยชน์ต่อธุรกิจในปัจจุบันหรือไม่ อย่างไร

_____

_____

_____

3. บริษัทของท่านมีการใช้งาน micro-blog เพื่ออะไรบ้าง

_____

_____

_____

_____

4. บริษัทของท่านมีทีมงานเฉพาะที่ทำหน้าที่รับผิดชอบการเขียนหรือการอ่านเนื้อหาใน micro-blog หรือไม่ ถ้ามีเจ้าหน้าที่เหล่านั้นต้องทำอะไรบ้าง

_____

_____

_____

_____

5. ท่านมีส่วนกระตุ้นหรือห้ามให้พนักงานภายในบริษัทใช้ micro-blog หรือไม่ อย่างไร

_____

_____

_____

_____

วัตถุประสงค์ 2 เพื่อทำความเข้าใจการใช้งาน micro-blog กับการวิจัยตลาด (40 นาที)

6. บริษัทของท่านมีการทำวิจัยการตลาดด้วยวิธีการใดบ้าง ท่านให้ความเชื่อถือวิธีการเหล่านั้นเพียงใด เช่น การวิจัยเชิงคุณภาพ (interview, observation, focus group discussion) การวิจัยเชิงปริมาณ (market survey)

_____

_____

_____

_____

_____

_____

_____

7. หากท่านสามารถนำ micro-blog ไปใช้ในการศึกษาวิจัยการตลาด ท่านคิดว่ามีประเด็นข้อจำกัด ปัญหา และอุปสรรค อะไรบ้าง

_____

_____

_____

_____

_____

_____

_____

วัตถุประสงค์ 3 เพื่อหาความต้องการที่ชัดเจนไปใช้ในการพัฒนาซอฟต์แวร์ต้นแบบ (50 นาที)

8. ธุรกิจของท่านมีการแนะนำ สินค้า/บริการ ใหม่บ่อยครั้งมากน้อยแค่ไหน

 ทุกสัปดาห์ ทุกเดือน ทุกไตรมาส (3 เดือน) ทุกปี

9. ท่านคิดว่าปัจจุบันใน micro-blog มีการพูดถึง สินค้า/บริการ ของท่าน หรือในอุตสาหกรรมของท่านมากน้อยเพียงใด

_____

_____

_____

_____

10. ท่านคิดว่าการวิเคราะห์ feedback เกี่ยวกับ สินค้า/บริการ ของท่านจาก micro-blog จะเป็นประโยชน์มากในช่วงเวลาใด กรุณาเรียงลำดับความสำคัญ 1 น้อยที่สุด 5 มากที่สุด

- การพัฒนาผลิตภัณฑ์ใหม่ (R&D)
- การติดตามความพึงพอใจสำหรับสินค้า/บริการในปัจจุบัน (CRM)
- จับตาคู่แข่งในตลาด (Competitive advantage)
- ปรับกลยุทธทางการตลาด (4P marketing)
- อื่น ๆ โปรดระบุ_____

11. ถ้าระบบสามารถบ่งบอกถึงความคิดเห็นว่าเป็นแง่บวกหรือลบของฟีเจอร์ของสินค้า/บริการของท่าน ท่านจะนำไปใช้ประโยชน์ต่อได้อย่างไร

_____

_____

_____

_____

_____

_____

วัตถุประสงค์ 4 เพื่อประเมินความเป็นไปได้ทางธุรกิจของบริการวิจัยการตลาดบน micro-blog (20 นาที)

12. ธุรกิจของท่านมีการใช้งบประมาณอย่างไรในการทำการวิจัยตลาด

- _____% ของยอดขาย
- _____% ของกำไร
- กำหนดคงตัว _____ บาทต่อปี
- อื่น ๆ โปรดระบุ _____

13. หากมีบริการวิจัยตลาดบน micro-blog ที่สามารถใช้งานได้ตลอดเวลา ท่านจะสนใจใช้บริการนี้หรือไม่อย่างไร

14. ท่านคิดว่าธุรกิจประเภทไหน เหมาะสมกับการวิจัยการตลาดบน micro-blog

15. อะไรคือสิ่งที่ท่านคาดหวังจากบริการการวิจัยการตลาดบน micro-blog มากที่สุด

APPENDIX B

กำหนดการสนทนากลุ่ม (Focus group discussion)
หัวข้อ “Innovative Marketing Tool by Applying Opinion Mining on Micro-blogs”
โดย นายวิลาส ฉ่ำเลิศวัฒน์ ในความดูแลที่ปรึกษาโดย ผศ.ดร.ภัทรสินี ภัทรโกศล และ ดร.ทิพภากร รังคสิริ
วันที่ 17 กุมภาพันธ์ 2555 เวลา 13:30 – 16:00 น.
ณ. ห้องประชุม 3 หลักสูตรธุรกิจเทคโนโลยีและการจัดการนวัตกรรม จุฬาลงกรณ์มหาวิทยาลัย
ชั้น 14 อาคารจัตุรัสจามจุรี

13:30 – 14:00 ลงทะเบียน

14:00 – 14:15 กล่าวต้อนรับ แนะนำผู้ควบคุมสนทนากลุ่ม ดร.ทิพภากร รังคสิริ
แนะนำผู้เข้าร่วมสนทนากลุ่ม อันประกอบด้วย

ผศ.ม.ล. สาวิกา อุณหนันท์
ผู้เชี่ยวชาญด้านการตลาด จุฬาลงกรณ์มหาวิทยาลัย

ดร. ชัชวิทย์ อาภรณ์เทวัญ
ผู้เชี่ยวชาญด้านเทคโนโลยี จุฬาลงกรณ์มหาวิทยาลัย

ภูมิใจ กฤติยานนท์
ผู้เชี่ยวชาญด้าน Smartphone บริษัท แอดวานซ์ อินโฟร์ เซอร์วิส จำกัด (มหาชน)

ศวิศิษฐ์ เพื่อวงษ์
ผู้เชี่ยวชาญด้าน Smartphone บริษัท ดิจิตอล มีเดีย แอดเวอร์ไทซิ่ง จำกัด

ธีรวัฒน์ สุวรรณพินิจ
ผู้เชี่ยวชาญด้าน Smartphone บริษัท เอ็ม วิชั่น จำกัด

จิรัสย์ พิพัฒน์วศิน
ผู้เชี่ยวชาญด้าน Smartphone บริษัท โบลด์ บิสเน็ท จำกัด

วรวิสุทธิ์ ภิญโญยาง
ผู้เชี่ยวชาญด้าน Social media บริษัท ดิจิบ๊อกซ์ จำกัด

พัฐวร ผ่องแผ้ว
ผู้เชี่ยวชาญด้าน Social media บริษัท ดีม๊อก จำกัด

พเนิน อัศววิภาส
ผู้เชี่ยวชาญด้าน Social media บริษัท วันบิต แมทเทอร์ จำกัด

14:15 – 14:30 แนะนำวัตถุประสงค์การสนทนากลุ่ม พร้อมนำเสนอภาพรวมของงานวิจัย

14:30 – 14:40 ส่วนที่ 1: บทสัมภาษณ์เกี่ยวกับตัวผู้เชี่ยวชาญ

14:40 – 14:55 ส่วนที่ 2 : การประเมินแนวคิดต้นแบบ Micro-blog Sentiment Analysis System

14:55 – 15:10 ส่วนที่ 3 : การยอมรับในต้นแบบนวัตกรรมนี้

15:10 – 15:50 ส่วนที่ 4 : คำแนะนำเพิ่มเติมสำหรับต้นแบบนวัตกรรมนี้

15:50 – 16:00 สรุปการอภิปราย ซักถามข้อเสนอแนะทั่วไป กล่าวขอบคุณ พร้อมมอบของที่ระลึก

แบบสอบถามผู้เชี่ยวชาญเพื่อประเมินผลงานวิจัย
เป็นส่วนหนึ่งของงาน ดุษฎีนิพนธ์ นายวิลาส ฉ่ำเลิศวัฒน์

การสอบถามในครั้งนี้มีวัตถุประสงค์**เพื่อประเมินการยอมรับต้นแบบ Micro-blog Sentiment Analysis System (MSAS)** จากท่านผู้เชี่ยวชาญที่เกี่ยวข้องกับงานวิจัยรวมถึงให้คำแนะนำและข้อเสนอแนะในการยอมรับและปรับปรุงนวัตกรรมให้มีความสมบูรณ์ก่อนนำไปสู่เชิงพาณิชย์

แบบสอบถามมีทั้งหมด 6 หน้าประกอบไปด้วย 4 ส่วนดังนี้

ส่วนที่ 1 : บทสัมภาษณ์เกี่ยวกับตัวผู้เชี่ยวชาญ

ส่วนที่ 2 : การประเมินแนวคิดต้นแบบ Micro-blog Sentiment Analysis System

ส่วนที่ 3 : การยอมรับต้นแบบนวัตกรรมนี้

ส่วนที่ 4 : คำแนะนำเพิ่มเติมสำหรับต้นแบบนวัตกรรมนี้

ดังนั้นจึงขอความอนุเคราะห์จากท่านผู้เชี่ยวชาญทุกท่านทำเครื่องหมาย✓หรือเติมข้อความลงในช่องว่างที่ตรงกับความจริงโดยข้อมูลที่ได้จะนำผลไปใช้ในทางการศึกษาต่อไปขอขอบคุณทุกท่านเป็นอย่างยิ่งที่กรุณาสละเวลาที่ให้ความร่วมมือในการตอบแบบสอบถามครั้งนี้

คำชี้แจ้ง : Sentiment Analysis เป็นกระบวนการวิเคราะห์ความรู้สึกจากตัวอักษร โดยผู้วิจัยได้รวบรวมข้อความจาก Twitter ในช่วงเดือน มี.ค. 53 - มิ.ย. 53 มาจำนวน 100,000 tweets เพื่อใช้ในการทดลองครั้งนี้ ท่านสามารถดูผลการทดลองเพิ่มเติมได้จาก http://wilas.chamlertwat.in.th/twittersentiment

วัตถุประสงค์

1. เพื่อสอบถามความคิดเห็นในแง่มุมต่างๆ เกี่ยวกับนวัตกรรมนี้
2. เพื่อประเมินความถูกต้องเหมาะสมของนวัตกรรมนี้ในอุตสาหกรรม smartphone
3. เพื่อประเมินความเป็นไปได้เชิงพาณิชย์สำหรับนวัตกรรมนี้

<u>ส่วนที่ 1</u>: บทสัมภาษณ์เกี่ยวกับตัวผู้เชี่ยวชาญ

ชื่อ-นามสกุล (ผู้เชี่ยวชาญ) .....
ตำแหน่ง.....
องค์กร/หน่วยงาน.....
ความเชี่ยวชาญ (เลือกได้มากกว่า 1 ข้อ)
☐Twitter ☐ Smartphone ☐ Consumer insight

คำถาม

1. ท่านใช้ Twitter ไหม บ่อยแค่ไหน
   ☐ ไม่ค่อยได้ใช้ ☐ ไม่บ่อย แต่ก็มีบ้าง ☐ ตลอดเวลาที่นึกได้
2. ท่านเคยใช้ search.twitter.com บ่อยแค่ไหน
   ☐ ไม่เคยใช้ ☐ นานๆ ที ☐ ใช้เป็นประจำ
3. ท่านคิดว่าข้อมูลที่มาจาก Twitter มีประโยชน์ไหม
   ☐ ไม่มีประโยชน์ เต็มไม่ด้วยขยะ
   ☐ มีประโยชน์บางแต่ต้องกรองก่อน
   ☐ มีประโยชน์มาก สามารถนำไปประกอบการตัดสินใจในเรื่องต่างๆ ได้
4. ปัจจุบันท่านใช้มือถือยี่ห้ออะไร รุ่นไหน
   .....
5. กรุณาให้คะแนนความพึงพอใจ มือถือของท่านในด้านต่างๆ ต่อไปนี้

ฟีเจอร์ด้านต่างๆ	**ความพึงพอใจในโทรศัพท์มือถือของท่าน (คะแนน 1 น้อยสุด ถึง 5 มากสุด)**
Screen (หน้าจอ)	
Application (โปรแกรมแอพพลิเคชั่นต่างๆ)	
Network (เครือข่ายการรับสัญญาณ)	
System (ระบบโดยรวม เช่น processor)	
Camera (กล้อง)	
Capacity (ความจุ)	
Power (ระบบไฟ)	
Sensor (ระบบ sensor)	
Accessories (อุปกรณ์เสริมต่างๆ)	
Size (ขนาด)	

6. ท่านเปลี่ยนโทรศัพท์มือถือบ่อยแค่ไหน
   ☐ เป็นปีๆ กว่าจะเปลี่ยน ☐ ประมาณไม่เกินหนึ่งปี ☐ แทบทุกครั้งที่มีโทรศัพท์รุ่นใหม่ออกมา
7. ท่านชื่นชอบ หรืออยากได้มือถือรุ่นใดเป็นพิเศษหรือไม่
   .....
8. ท่านติดตามความเคลื่อนไหวของรุ่นโทรศัพท์มือถือบ่อยแค่ไหน
   ☐ ไม่ได้ติดตามเท่าไร ☐ เป็นระยะๆ แล้วแต่ช่วงอยากเปลี่ยนมือถือใหม่ ☐ ตลอดเวลา เป็นงานอดิเรก

9. ท่านคิดว่าพฤติกรรมต่อไปนี้ ข้อใดเป็นวิธีการที่ดีที่สุดในการเข้าใจความต้องการของลูกค้า
   - ☐ **ทำแบบสอบถาม/ Focus group**
   - ☐ **สังเกตจากพฤติกรรมของผู้ใช้**
   - ☐ **ทำนายด้วยความชำนาญของทีม R&D ภายในองค์กร**

<u>ส่วนที่ 2</u> : การประเมินแนวคิดต้นแบบ Micro-blog Sentiment Analysis System

1. ท่านเคยรู้จักระบบ Sentiment Analysis มาก่อนหรือไม่

   ☐ **ไม่เคย** ☐ **เคย โปรดระบุชื่อโปรแกรมหรือเว็บไซต์**

   .....

   .....

   .....

2. ท่านคิดว่าการนำเอาความรู้สึกของผู้ใช้งานมาเป็นปัจจัยหรือเกณฑ์สำคัญในการพัฒนาผลิตภัณฑ์ จะช่วยให้การพัฒนาผลิตภัณฑ์ในรุ่นถัดไปตรงใจผู้ใช้มากขึ้นได้ ใช่หรือไม่

   ☐ ใช่ ☐ ไม่ใช่ (โปรดระบุความคิดเห็นของท่าน)

   .....

   .....

   .....

3. ตารางด้านล่างนี้เป็นผลจาก Micro-blog Sentiment Analysis System จาก Twitter โดยช่อง + คือจำนวนคนที่ให้ความเห็นเชิง Positive กับฟีเจอร์นั้น และ – คือจำนวนคนที่ให้ความเห็นเชิง Negative กับฟีเจอร์นั้นๆ

	Blackberry		HTC		iPhone		Motorola	
	+	-	+	-	+	-	+	-
Screen	1	0	1	0	4	1	0	0
Application	150	106	43	20	1071	473	3	3
Network	20	21	52	15	104	69	18	12
System	0	0	21	1	7	1	12	15
Camera	66	60	138	25	176	73	36	7
Capacity	2	4	1	1	4	3	0	1
Power	53	51	65	39	101	123	35	12
Sensor	1	0	0	0	0	0	0	0
Accessories	9	2	3	0	6	14	2	1
Size	2	0	3	0	0	0	0	0

	Nexus		Nokia		Palm		Samsung	
	+	-	+	-	+	-	+	-
Screen	0	1	0	0	0	0	7	0
Application	37	12	2	4	7	1	10	10
Network	33	14	17	5	1	0	3	2
System	4	0	0	0	0	0	0	0
Camera	175	139	6	3	2	0	62	7
Capacity	1	0	1	1	0	0	0	1
Power	16	19	4	0	4	0	4	1
Sensor	1	0	0	0	0	0	0	0
Accessories	1	3	5	0	1	0	0	2
Size	0	2	0	0	0	0	0	0

จากตารางข้างต้นท่านเห็นด้วยหรือไม่เห็นด้วยกับผลลัพธ์จากการทดลองสำหรับมือถือยี่ห้อใดบ้าง เพราะเหตุใด

	เห็นด้วย	ไม่เห็นด้วย	อธิบายเพิ่มเติม
Blackberry			
HTC			
iPhone			
Motorola			
Nexus			
Nokia			
Palm			
Samsung			

<u>ส่วนที่ 3</u> : การยอมรับในต้นแบบนวัตกรรมนี้

ปัจจัยการยอมรับนวัตกรรม	ระดับความคิดเห็น				
	มากที่สุด (5)	มาก (4)	ปานกลาง (3)	น้อย (2)	น้อยที่สุด (1)
การรับรู้ประโยชน์ของนวัตกรรม					
ช่วยประหยัดเวลาในการตรวจสอบความต้องการของผู้บริโภค					
เพิ่มประสิทธิภาพและความถูกต้องในการเข้าใจความต้องการของผู้บริโภค					
ทำให้ทราบข้อมูลใหม่ๆ ที่ไม่เคยรู้มาก่อน					
การใช้งานง่ายของนวัตกรรม					
นวัตกรรมนี้มีการนำเสนอที่ง่ายต่อการเข้าใจ					
องค์กรสามารถนำผลลัพธ์ไปใช้ต่อได้สะดวก					
ช่วยลดความยุ่งยากและขั้นตอนในการทำวิจัยเกี่ยวกับผู้บริโภค					
ทัศนคติการใช้งาน					
ท่านสนใจที่จะใช้นวัตกรรมนี้ในอนาคต	☐ ใช่		☐ ไม่ใช่		
นวัตกรรมนี้ช่วยแนะนำทิศทางในการผลิตผลิตภัณฑ์ใหม่ได้จริง	☐ ใช่		☐ ไม่ใช่		
นวัตกรรมนี้มีความใหม่ในการนำไปใช้กับอุตสาหกรรมของท่าน	☐ ใช่		☐ ไม่ใช่		

<u>ส่วนที่ 4</u> : คำแนะนำเพิ่มเติมสำหรับต้นแบบนวัตกรรมนี้

1. **ท่านชื่นชอบส่วนไหนมากที่สุดสำหรับการออกแบบต้นแบบนวัตกรรมนี้พร้อมเหตุผลประกอบ**

   .....

   .....

   .....

   .....

2. **ส่วนใดที่ท่านไม่ชอบ**และคิดว่าไม่ควรมีในการออกแบบต้นแบบนวัตกรรมนี้เลยพร้อมเหตุผลประกอบ

   .....

   .....

   .....

   .....

3. **ท่านมีข้อเสนอแนะเพิ่มเติมและแนวทางในการปรับปรุงเกี่ยวกับการพัฒนาต้นแบบนวัตกรรมในครั้งนี้หรือไม่**

   .....

   .....

   .....

   .....

4. **ท่านคิดว่าคู่แข่งสำหรับนวัตกรรมในรูปแบบนี้มีมากน้อยเพียงใดและท่านมองว่าข้อได้เปรียบของนวัตกรรมนี้จะช่วยในการแข่งขันได้หรือไม่**

   .....

   .....

   .....

   .....

5. **ท่านคิดว่านวัตกรรมนี้มีความเป็นไปทางธุรกิจมากน้อยเพียงใด และควรมีรูปแบบการคิดค่าบริการอย่างไร**

   .....

   .....

   .....

   .....

6. **ท่านมีคำแนะนำอื่นๆ อีกหรือไม่**

   .....

   .....

   .....

   .....

ขอขอบพระคุณทุกท่านที่สละเวลาให้ความร่วมมือในการตอบแบบสอบถาม
ทางผู้วิจัยใคร่ขอกราบขอบพระคุณทุกท่านมาณโอกาสนี้

BIOGRAPHY

Mr. Wilas Chamlertwat is an ordinary man who tries to pursue his dream on Technopreneur way. His expertise surround on web development and social media. In his leisure time, the best activity is playing with his daughter and giving time to his family.

Education:

Jun 2007 - May 2012 Doctor of Philosophy Program

Technopreneurship and Innovation Management (Interdisciplinary Program)

Graduate School, Chulalongkorn University

Aug 2002 - May 2004 Master of Business Administration

Concentration on Computer Information Systems

San Francisco State University, San Francisco, CA, U.S.A.

May 1995 - Mar 1999 Bachelor of Science in Computer Science (2^{nd} Honor)

Mathematics Department

Chulalongkorn University

Working experiences:

Jan 2012 - present Managing Director at ServePro Co., Ltd.

Jul 2011 - present Chief Technology Officer at BOXBOX Co., Ltd.

Jul 2007 - present Chief Strategy Officer at DotArai Co., Ltd.

Jan 2006 - present Managing Director at WoNe' Co., Ltd.

Mar 2005 - present Strategic planner at T.H.NIC Co., Ltd.